M 2016

U.PORTO

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

METHYLOME OF BREAST CANCER MOLECULAR SUBTYPES:

DIAGNOSIS AND PROGNOSIS BIOMARKERS

SOFIA RAQUEL FERNANDES SALTA
DISSERTAÇÃO DE MESTRADO APRESENTADA
AO INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
DA UNIVERSIDADE DO PORTO EM
ONCOLOGIA – ESPECIALIZAÇÃO EM ONCOLOGIA MOLECULAR

Sofia Raquel Fernandes Salta

Methylome of Breast Cancer Molecular Subtypes: Diagnosis and Prognosis Biomarkers

Dissertação de Candidatura ao grau de Mestre em Oncologia – Especialização em Oncologia Molecular submetida ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto.

Orientadora

Professora Doutora Carmen de Lurdes Fonseca Jerónimo

Professora Associada convidada com Agregação

Departamento de Patologia e Imunologia Molecular

Instituto de Ciências Biomédicas Abel Salazar

Universidade do Porto

Investigadora Auxiliar e Coordenadora do Grupo de Epigenética e Biologia do Cancro

Centro de Investigação

Instituto Português de Oncologia – Porto

Coorientador

Professor Doutor Rui Manuel Ferreira Henrique

Professor Catedrático Convidado

Departamento de Patologia e Imunologia Molecular

Instituto de Ciências Biomédicas Abel Salazar

Universidade do Porto

Diretor do Serviço de Anatomia Patológica

Investigador Sénior do Grupo de Epigenética e Biologia do Cancro

Centro de Investigação

Instituto Português de Oncologia – Porto

“Success is not final, failure is not fatal: it is the courage to continue that counts”.

Winston Churchill

Financial support

This study was funded by a grant of the Research Centre of Portuguese Oncology Institute of Porto (CI-IPOP-74-2016).

ACKNOWLEDGEMENTS

A realização desta tese não seria possível sem a ajuda e colaboração de diversas pessoas às quais expresso os mesmos agradecimentos.

Em primeiro lugar, à minha orientadora, a Professora Doutora Carmen Jerónimo e ao meu co-orientador, o Professor Doutor Rui Henrique expresso os meus maiores agradecimentos, por me permitirem integrar o grupo de Epigenética e Biologia do Cancro. Agradeço o voto de confiança para a realização desta tese, a partilha de conhecimentos, o apoio, a disponibilidade e a possibilidade de desenvolver novas capacidades na área de investigação em Oncologia.

Ao Professor Doutor Manuel Teixeira, na qualidade de Diretor do Centro de Investigação do IPO do Porto, agradeço a possibilidade da realização da minha tese neste centro de investigação.

À Professora Doutora Berta Martins, na qualidade de Diretora do Mestrado em Oncologia, agradeço a aceitação e seleção da minha candidatura para este mestrado.

Ao Serviço Anatomia Patológica do IPO do Porto em particular à Drª Margarida Caldas, pela ajuda na seleção e visualização dos casos e pela disponibilidade para me esclarecer as minhas dúvidas, à Técnica Paula Lopes, pela realização de toda a imuno-histoquímica para a caracterização da série e pela ajuda sempre que necessitei e à Técnica Isa Carneiro pela ajuda indispensável para aprender a realizar os cortes no criostato.

Ao Serviço de Epidemiologia do IPO do Porto na pessoa do Engenheiro Luís Antunes por toda a ajuda na realização da análise estatística que permitiu as conclusões deste estudo.

Ao Serviço de Patologia da Mama desta instituição em particular à Drª Susana Sousa e ao Dr. Mário Fontes e Sousa agradeço todo o tempo disponibilizado na recolha de informação clínica dos doentes estudados e no esclarecimento das minhas dúvidas. Agradeço ainda ao Dr. Fernando Castro, ao Dr. Pedro Antunes e a toda a equipa de enfermagem (em especial Enfª Elisabete, Enfª Elvira, Enfª Amelia, Enfª Maria do Céu, Enfª Idalina e Enfª Isabel) desta clínica pela disponibilidade e colaboração para este trabalho.

Agradeço aos membros do Grupo de Epigenética e Biologia do Cancro, uma equipa que me acolheu e que permitiu sem dúvida o desenvolvimento e concretização desta tese. À Ana Luís agradeço a ajuda constante ao longo deste trabalho, que desde aprendizagem das técnicas utilizadas neste trabalho à análise estatística, nunca esquecendo todo o apoio ao longo destes meses. À Inês e ao João agradeço o espirito crítico e ajuda na resolução de diversos problemas que, inevitavelmente, foram surgindo ao longo deste trabalho. À Francisca, Maria João e Eva, as minhas colegas de mestrado, agradeço todo o apoio ao

longo destes meses constituídos por momentos tantos momentos desde das horas de almoço à discussão dos resultados. A vocês muito obrigada. À Micaela agradeço as horas a discutir o nosso trabalho e pelo apoio. Agradeço ainda a todos os outros membros do Grupo de Epigenética e Biologia do Cancro, atuais e passados, e de alguma forma quer pelo apoio quer pela ajuda contribuíram para que este trabalho fosso possível. A todos muito obrigada.

Agradeço ainda à D. Luísa a sua constante disponibilidade e ajuda fundamentais ao trabalho no laboratório.

Por fim, mas não menos importante, agradeço aos meus pais e à minha irmã, pessoas sem as quais todo o trabalho e participação no mestrado em Oncologia não seria possível. A vós agradeço todo o apoio e compreensão para que o trabalho que apresento aqui hoje foi possível.

RESUMO

O cancro da mama é o segundo cancro mais incidente a nível mundial e representa a quinta causa de morte por cancro. Em Portugal é o cancro mais incidente e a principal causa de morte por cancro no sexo feminino. Apesar dos programas de rastreio baseados na mamografia, implementados sobretudo nos países mais desenvolvidos, permitirem o diagnóstico de muitos cancros num estadio inicial, levando a uma diminuição da taxa de mortalidade por cancro da mama nestes países, esta é ainda uma das mais elevadas. Devido à grande heterogeneidade do cancro da mama, múltiplos estudos têm sido realizados de forma a melhor caracterizar esta doença e, consequentemente, melhor estratificar os doentes. Foram descritos quatro subtipos moleculares principais no cancro da mama que apresentam diferente valor prognóstico. Contudo, a sua caracterização na prática clínica apresenta, ainda, limitações. A descoberta do papel da epigenética, nomeadamente da metilação do DNA, na regulação da normal expressão génica e o seu papel na carcinogénese de diversas neoplasias, é responsável pelas inúmeras publicações nesta área de investigação nos últimos anos. A metilação aberrante de diversos genes associados ao cancro da mama está largamente descrita, tendo sido proposta como biomarcador de diagnóstico e prognóstico.

Este trabalho teve como principal objetivo verificar se a metilação de um painel de seis genes (*APC*, *BRCA1*, *CCND2*, *FOXA1*, *RASSF1A* e *SCGB3A1*) possui potencial para ser utilizado na detecção/diagnóstico dos diversos subtipos moleculares de cancro da mama.

Os níveis de metilação do promotor dos seis genes foram avaliados por PCR quantitativo específico para metilação em DNA modificado por bissulfito de sódio em 137 amostras de tecido de cancro da mama e 18 amostras de mama normal. Para a determinação do desempenho diagnóstico da metilação destes genes foi realizada a análise de curva ROC. Os níveis de metilação foram ainda associados com os subtipos moleculares (previamente determinados por imuno-histoquímica) e com as variáveis clinico-patológicas relevantes. A análise de sobrevivência livre de doença e específica de doença foi realizada segundo o método de Kaplan-Meyer, utilizando o teste de log-rank e a regressão de Cox para comparações e tendo-se avaliado o valor prognóstico da metilação dos genes estudados.

Os nossos resultados permitiram confirmar o potencial diagnóstico da metilação do DNA para cinco dos seis genes estudados. A utilização de um painel baseado na metilação de três genes (*APC, FOXA1* e *RASSF1A*) permitiu a descriminação de amostras neoplásicas de amostras normais com uma especificidade, sensibilidade e acuidade superiores a 94%. Não foi encontrada nenhuma associação entre a metilação dos seis

genes e qualquer subtipo molecular. Contudo, os níveis de metilação do promotor do *APC* e do *SCGB3A1* correlacionaram-se com os subtipos histológicos (nomeadamente os carcinomas de subtipos especiais) e a metilação do *FOXA1* e do *RASSF1A* com o *status* dos recetores hormonais. Na nossa série níveis elevados de metilação do *FOXA1* associaram-se com uma pior sobrevivência específica de doença.

Este trabalho confirmou o potencial do uso da metilação do DNA como biomarcador de diagnóstico com um painel com alta sensibilidade e especificidade capaz de detetar qualquer subtipo molecular de cancro da mama. Adicionalmente, os resultados revelaram o potencial da metilação do *FOXA1* como fator prognóstico, até à data não reportado na literatura.

Este trabalho permitiu, assim, revelar novos potenciais marcadores de diagnóstico e prognóstico para o cancro de mama cuja validação em estudos prospetivos poderá revelar grande utilidade clínica.

ABSTRACT

Breast cancer is the second most incident cancer worldwide and the fifth cause of cancer-related death. In Portugal, it is the most incident cancer and the leading cause of cancer death among women. Although mammography-based screening implementation in developed countries allowed for mortality rate decrease, it is still very high. Breast cancer is a heterogeneous disease and multiples studies have attempted to better understand this disease and, consequently, better stratify breast cancer patients. Four major molecular subtypes are recognized in breast cancer, which have different prognosis. However, the use of molecular characterization in clinical practice is still limited. Epigenetics mechanisms, such as DNA methylation, are important gene expression regulators in normal cells and its deregulation in cancer has been increasingly recognized. Aberrant DNA methylation in breast cancer has been widely described and several studies have unveiled its diagnostic and prognostic potential.

The main goal of this study was to determine the diagnostic and prognostic potential of promoter methylation levels of six genes (*APC, BRCA1, CCND2, FOXA1, RASSF1A* and *SCGB3A1*), as well as its ability to discriminate among the breast cancer molecular subtypes.

The methylation levels of all candidate genes were assessed by quantitative methylation-specific PCR using sodium-bisulphite modified DNA from fresh frozen samples of breast cancer (n=137) and normal breast (n=18). To assess the diagnostic potential, ROC curve analysis was performed. Promoters' methylation levels were correlated with molecular subtypes (previously determined by immunohistochemistry) and with standard clinicopathological parameters. To verify the prognostic potential of methylation levels, log-rank test and Cox regression was performed for disease-free and disease-specific survival.

Five out of the six candidate genes demonstrated diagnostic potential. A three-gene methylation panel (*APC, FOXA1* and *SCGB3A1*) could discriminate normal from cancerous breast tissue with specificity, sensitivity and accuracy above 94%. Methylation levels of these six genes did not correlate with any molecular subtype of breast cancer. Nevertheless, *APC* and *SCGB3A1* methylation levels associated with histological subtypes (specifically with special subtype carcinomas) whereas *FOXA1* and *RASSF1A* methylation levels associated with hormone receptor status. In our series, high *FOXA1* methylation levels associated with shorter disease-specific survival.

Thus, we confirmed the diagnostic potential of DNA methylation levels in breast cancer. Importantly, the three-gene panel displayed high specificity, sensitivity and accuracy for breast cancer detection regardless of molecular subtype. Moreover, the results uncovered the prognostic potential of *FOXA1* promoter methylation.

This work revealed new potential diagnostic and prognostic biomarkers for breast cancer and its validation in large prospective studies is mandatory to evaluate its clinical usefulness.

TABLE OF CONTENTS

FIGURES INDEX

TABLES INDEX

LIST OF ABBREVIATIONS

µg – microgram

µL – microliters

µm – micrometers

ACTβ – Actin β

AJCC – American Joint Committee on Cancer

APC – Adenomatosis polyposis coli

ASCO – American Society of Clinical Oncology

AUC – Area Under the Curve

BrC – Breast Cancer

BRCA1 – BRCA1, DNA repair associated

BRCA2 – BRCA2, DNA repair associated

CCND2 – Cyclin D2

cfDNA – cell-free DNA

CNVs – Copy number variations

CpG - Cytosine-phosphate-Guanine

DCIS – Ductal Carcinoma in situ

DNA – Deoxyribonucleic Acid

DNMT – DNA methyltransferase

EDTA – Ethylenediamine teracetic acid

ER – Estrogen Receptor

ERBB2 – Erb-b2 receptor tyrosine kinase 2

ESMO – European Society for Medical Oncology

FNB – Fine-Needle Biopsy

FOXA1 – Fork-head box A1

HER2 – Human Epidermal Growth Factor 2 Receptor

HIN1 – High in Normal 1

HPF – High-power field

HRT – Hormone-replacement therapy

IDC – Invasive Ductal Carcinoma

IHC – Immunohistochemistry

ILC – Invasive Lobular Carcinoma

IQR – Interquartile Range

LCIS – Lobular Carcinoma in situ

M – Molar

mg – Milligram

miRNA – microRNA

mL – milliliters

mM – Milimolar

mRNA – messenger RNA

MSP – Methylation Specific Methylation

MTC – Mixed Type Carcinoma

NaCl – Sodium chlorine

NBr – Normal Breast

NCB – Needle-Core Biopsy

ncRNAs – non-coding RNAs

NPV – Negative Predictive Value

NST – No Special Type

PCR – Polymerase Chain Reaction

PPV – Positive Predictive Value

PR – Progesterone Receptor

PTEN – Phosphatase and tensin homolog

qMSP – Quantitative Specific Methylation PCR

RASSF1A – Ras association domain family 1 isoform A

RNA – Ribonucleic Acid

ROC – Receiver Operating Characteristic

rpm – rotation per minute

SCGB3A1 – Secretoglobin family 3A member 1

SDS – Sodium Dodecyl Sulfate

SE – Buffer solution

SEER – Surveillance, Epidemiology, and End Results

SNPs – Single Nucleotide Polymorphisms

SSC – Special Subtypes Carcinoma

TNBC – Triple Negative Breast Cancer

TP53 – Tumor protein p53

UICC – Union for International Cancer Control

INTRODUCTION

BREAST CANCER

EPIDEMIOLOGY

Breast cancer (BrC) is the 2nd cancer more incident worldwide, corresponding about to 11.9% of all cancer incidence. Within an estimated 1.67 million new cases in 2012 is the most common cancer in women (about 25% of all cancer). The BrC incidence is higher in more developed regions, such Northern America, Australia, and Northern and Western Europe (Figure 1)[1]. These differences are mainly due to screening programs that allow the detection of early invasive cancers in the former countries [2].

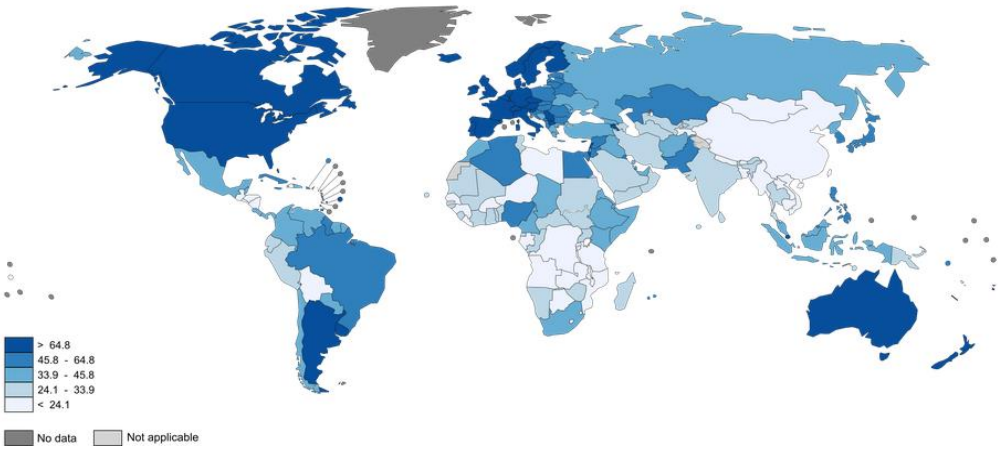

Figure 1 – Estimated Age-Standardized Breast Cancer Incidence Worlwide in 2012[3].

Despite the high incidence of BrC is the 5th cause of death by cancer due to the access of health care and the early cancer detection. Worldwide in 2012 522 thousand deaths by BrC were estimated [1, 2].

In Europe, BrC was the leading cancer in both sexes (Figure 2). In 2012 were estimated about 484 (*per* 100 000) BrC new cases and 94.2 (*per* 100 000) deaths [4].

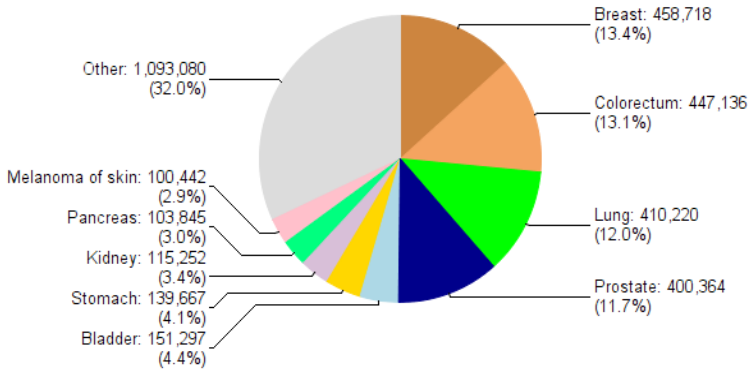

Figure 2 – Estimated number of cancer cases for both sexes in Europe in 2012 [3].

In Portugal, BrC was the leading cancer in 2012 with 67.6 (*per* 100 000) new cases and the 1st cause for cancer death in women with 13.1 (*per* 100 000) deaths (Figure 3)[4].

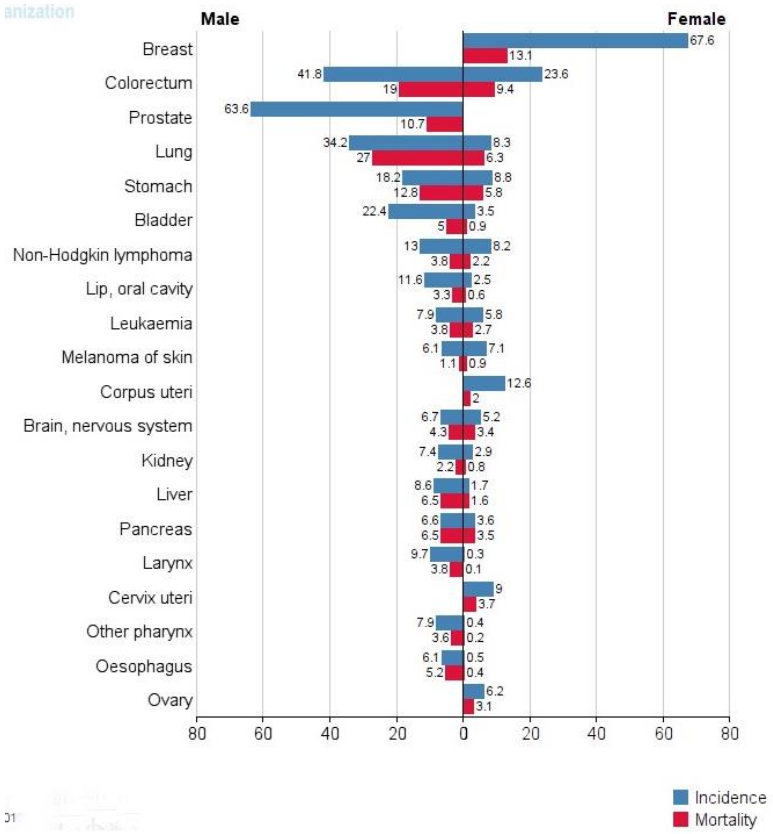

Figure 3 – Estimated Age-Standardized Incidence and Mortality Rates (per 100 000) in Portugal in 2012 [3].

This reality has turned BrC one of the main interest research field. Despite all the improvement in cancer early detection and treatment, BrC remains the foremost cause of cancer-related death in women due to development of recurrent and/or metastatic disease [1].

RISK FACTORS

The BrC is a complex disease with multiple factors associated. These factors include demographic, genetics, familiar factors and hormone exposure among others. Some of these factors are clustered in Table 1 by relative risk categories.

Table 1 – Magnitude of risk of known BrC risk factors. Adapted from [5, 6]

Relative Risk <2	Relative Risk 2-4	Relative Risk >4
• Early menarche • Late menopausal • Nulliparity • Hormone replacement therapy • Oral contracpetive use • Alcohol use Postmenopausal obesity	• One first-degree relative with BrC • Age > 30 years for first delivery • Proliferative breast disease	• *BRCA 1* or *BRCA2* Mutations • Atypical hyperplasia • Radiation exposure at early age

BrC - Breast Cancer

Demographic Factors

BrC incidence increases sharply with age, being commonly diagnosed in patients between 45 and 74 years (Figure 4)[7]. According to the Surveillance Epidemiology, and End Results (SEER) database, in the United States, the probability of a woman for developing BrC is 1 in 8 over a lifetime: 1 in 202 from birth to age 39 years of age, 1 in 26 from 40 to 59 years, and 1 in 28 from 60 to 69 years[8].

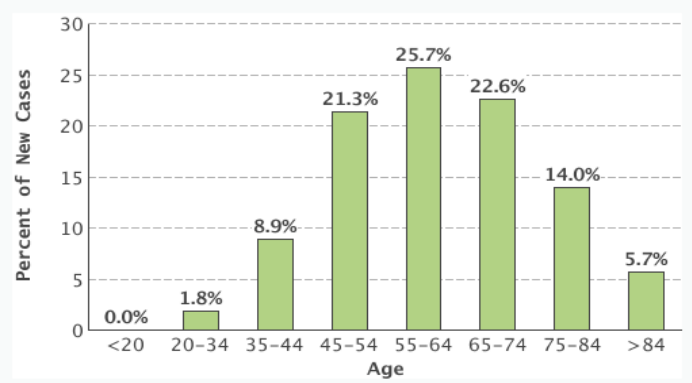

Figure 4 – Percent of new breast cancer cases by age group (Surveillance, Epidemiology, and End Results [SEER], 2009-2013) From [7].

Genetics and Familiar Factors

The family history of BrC is highly associated BrC risk. Women who have a first-degree relative with BrC history, have a higher risk of BrC. Moreover, this risk increases if the family member affected was diagnosed at young age (less than 40 years) [5, 6, 9].

Likewise, there are several syndromes associated with an increase in BrC risk. The most frequent hereditary mutations occur in *BRCA1, DNA repair associated (BRCA1)* and *BRCA*2 genes which are highly associated with BrC in younger age and premenopausal women. Less than 10% of all BrC cases are due to mutations in these genes, carriers of

BRCA1 and *BRCA2* mutations have an estimated lifetime risk of BrC between 26% and 85%. Additional genes such as *Tumor Protein p53 (TP53)*, *Phosphatase and Tensin homolog (PTEN)* and Lynch Syndrome associated genes are implicated in familiar BrC [5, 6, 9].

Reproductive Factors and Hormone Exposure

The risk of BrC is highly associated with women's reproductive history. Women with menarche at early age who remains nulliparous or with the first delivery at late age (30 years or later) have an increased risk of BrC development [5, 6, 9]. Late age at menopause is also risk factor to development of BrC [6, 9]. Indeed, these risk factors are associated with the exposure to elevated levels of estrogens, therefore, the surgical cessation of estrogen production by oophorectomy is protective against BrC development [6]. Moreover, the use of oral contraceptives or postmenopausal hormone-replacement therapy (HRT) associates with a slight but significant increase of BrC risk. The increased risk varies with usage duration and, in the case of HRT, if includes estrogen alone or estrogen plus progestin [5, 6, 9].

Dietary and Lifestyle Factors

The most emerging lifestyle factor for postmenopausal BrC is obesity. Indeed, an increased body mass index raises the risk to BrC due to elevation of endogenous estrogen concentrations in plasma. However, in premenopausal women, obesity seems have the oppose effect. Thus, the weight and its variation in lifetime seems to play a role in BrC risk [6, 9].

Moreover, a healthy diet with high intake of fruit and vegetables seems to decrease BrC risk, while high caloric diet and high red meat intake and saturated animal fat have the opposed effect [5, 9]. The consumption of alcohol is, as well, largely associated with an increased risk for BrC. These findings may be associated with endogenous hormone levels. Similarly, the active cigarrete smoking has been associated with an increase of BrC risk. However, the same is not described to passive smoking which only seem to impact in premenopausal BrC [6, 9].

Environmental Factors

The exposure to ionizing radiation increases BrC risk, moreover this increase arises with exposure in young age. Likewise, women who receive radiation to treatment of Hodgking lymphoma at young age have a higher risk to BrC [5, 6].

Benign Breast Disease

Benign Breast Disease is classified in two main categories: proliferative and non-proliferative breast lesions. Although non-proliferative lesions do not carry any BrC risk, the proliferative lesions are associated with a higher risk to BrC. Among these, atypical hyperplasia is the lesion with the higher risk to BrC, which significantly increases if associated with family history of BrC [5, 6].

DIAGNOSTIC, PROGNOSTIC FACTORS AND TREATMENT

Screening and Diagnosis

The diagnosis of BrC is a multidisciplinary task that requires several specialists in oncologic field such as surgeons, clinicians, pathologists and radiologists, among others.

The most common clinical presentation of BrC is a palpable mass. However, BrC may be presented clinically as persistent areas of pain or tenderness or nipple discharge [9, 10].

Due to the high incidence of BrC, eighteen European countries have implemented a national or regional population-based mammography screening program, in order to detect BrC at pre-clinical stage [11]. Typically, the screening programs plan to carry out a mammography every two years in women between 50 to 69 years old [11]. For women with high risk of BrC (e.g. women with *BRCA1* mutation) is recommended annual magnetic resonance imaging alongside or alternating with mammography every 6 months. High risk women should start the screening 10 years younger than the youngest case in the family [11].

The diagnosis of BrC includes the clinical examination in concomitance with imaging techniques, and pathological confirmation [11]. The pathological confirmation can be performed by fine-needle biopsy (FNB) or needle-core biopsy (NCB) with a good sensitivity for both techniques in palpable lesions. However, NBC presents higher sensitivity for detection of BrC in impalpable lesions, currently being the preferential method for pathological confirmation [9, 11]. The NCB is recommended for evaluation of FNA suspicious findings and evaluation of microcalcifications. When preoperative systemic therapies are planned, NCB is mandatory to ensure histopathological diagnosis and the assessment the established immunostaining markers [9, 11].

Histological Subtypes

BrC is a highly heterogenous disease and the current World Health Organization Classification of Tumours of Breast identifies more than 20 histological types.

The majority of the BrC have their origin at epithelial cells and can be subdivided into *in situ* and invasive carcinomas. The *in situ* carcinomas are defined as pre-invasive lesions in which neoplastic epithelial cells proliferate confined to the ductal/lobular tree of the breast without evidence of invasion through the basement membrane [12, 13]. *In situ* carcinomas are further subdivided in lobular carcinoma *in situ* (LCIS) e Ductal carcinoma *in situ* (DCIS). DCIS is more frequent than LCIS and these two lesions are distinguished by difference in the architectural and cytological features and not of the micro-anatomical site of origin (ducts or lobules) [9, 13]. The DCIS is an infrequent diagnosis in countries without mammographic screening programmes once it is present only in 2-3% of palpable BrC. With the implementation of mammographic screening programmes the DCIS represents 20-25% of newly diagnosed BrC. [9, 12].

Nonetheless, the invasive carcinomas are the most common BrC lesions (70-80% of malignant neoplasia of breast)[13]. The invasive carcinomas can generally be grouped in two categories: invasive carcinoma of no special type (NST) and special subtypes carcinomas (SSC). The invasive NST carcinoma and the invasive lobular carcinoma (ILC) represent the major types of BrC (Figure 5) [9, 12].

The invasive NST carcinoma is also known as ductal carcinoma NST or invasive ductal carcinoma (IDC) and represents up to 75% of all invasive carcinomas (Figure 5 A) [9]. This designation categorizes a heterogeneous group comprising the tumors that cannot be identified as any SSC and, consequently, morphological features vary among different tumors [9]. If a tumor 10-49% of a SSC pattern and the remaining are NST carcinoma it will be categorized as mixed type carcinoma (MTC) [9].

ILC is the most common of the SSC representing 5-15% of all BrC (Figure 5 B). Clinically most the women present an ill-defined palpable mass and the classic pattern is characterized by a proliferation of small cells, which lack cohesion and appear arranged in single-file linear cords that invade the stroma or individually dispersed trough a fibrous connective tissue [9].

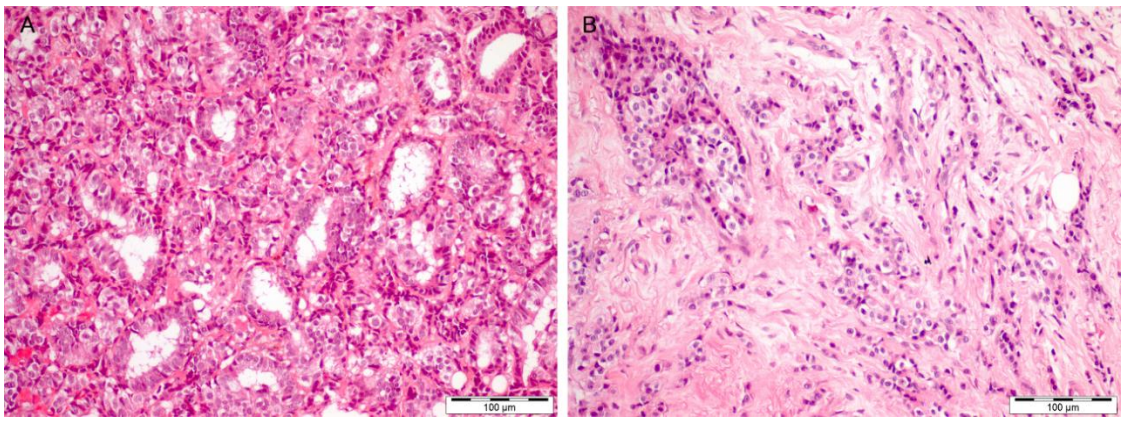

Figure 5 – Hematoxylin and eosin stained sections of the two most common histological subtypes of Breast Cancer, 200x. A – Ductal Invasive Carcinoma; B – Lobular Invasive Carcinoma.

SSC category includes also more than 10 others histological types which the frequency is much lower than the subtypes described above, such as mucinous carcinoma, carcinoma with medullary features, invasive papillary carcinoma and other.

Grading

All the breast tumors are graded based on assessment of tubule/gland formation, nuclear pleomorphism and mitotic counting. Currently, the Nottingham combined histologic grade (the Elston-Ellis modification of the Scarff-Bloom-Richardson grading system) is widely accepted [14] (Table 2). Grade is a powerful prognostic factor and it is a key component for clinical decision-making [9].

Table 2 – Nottingham combined histologic grade. Adapted from[14]

Criteria	**Score**		
	1	2	3
Glandular/Tubular differentiation	>75 % of tumor area forming glandular/ tubular structures	10-75 % of tumor area forming glandular/ tubular structures	<10 % of tumor area forming glandular/ tubular structures
Nuclear pleomorphism	Small nuclei with little increase in size in comparison with normal breast epithelial cells, regular outlines, uniform nuclear chromatin, little variation in size	Cells larger than normal with open vesicular nuclei, visible nucleoli, and moderate variability in both size and shape	Vesicular nuclei, often with prominent nucleoli, exhibiting marked variation in size and shape, occasionally with very large and bizarre forms
Mitotic Counting	≤4/10 HPF	5–9/10 HPF	≥10/10 HPF
	Overall Score		
	1	**2**	**3**
Total Score	3-5	6 or 7	8 or 9

HPF – High-power field

Staging and Prognosis

All the patients diagnosed with BrC are staged in order to evaluated disease extension and patient's prognosis. The most widely accepted system to stage BrC patients is TNM system published by the American Joint Committee on Cancer (AJCC)/ Union for International Cancer Control (UICC) [9]. This system combines information about the extent of cancer in primary site (tumor or **T**), the regional lymph node (nodes or **N**) and spread to the distant metastatic sites (metastases or **M**) [9, 15]. The combination of these 3

parameters allows for the establishment of five stages (0, I, II, III and IV) that summarize the information for making decisions concerning disease control and determine the value of systemic therapy (Appendix I). Moreover, stage is essential for comparison in clinical trials, epidemiological studies and others investigations [9].

Both clinical and pathological staging are used in patients with BrC. The clinical staging (c) (Appendix II) is determined using information identified prior to surgery or neoadjuvant therapy. It includes physical examination, with careful inspection and palpation of the skin, mammary gland, and lymph nodes, imaging (such as mammography and ultrasound), and pathologic examination of the breast or other tissues as appropriate to establish the diagnosis of breast carcinoma (e.g. FNB or NCB) [11, 15]. The pathologic staging (p) (Appendix II) includes information defined at surgery. It includes all data used for clinical staging, plus data from surgical exploration and resection as well as pathologic examination (gross and microscopic) of the primary carcinoma, regional lymph nodes, and metastatic sites (if available) [15].

Despite the information provided by TNM staging, for each BrC patient is also important to assess other biomarkers of prognostic outcome and predictive therapy response [11]. Currently, in the routine clinical management of BrC patients, estrogen receptor (ER) and progesterone receptor (PR) status is evaluated by immunohistochemistry (IHC) along with *Erb-b2 receptor tyrosine kinase 2 (ERBB2)* assessed by IHC and/or *in situ* hybridization.

Estrogen Receptor

ER is a nuclear transcription factor that stimulates the growth of normal breast epithelial cells when is active by hormone estrogen. ER is a crucial player in breast carcinogenesis, whose inhibition forms the mainstay of BrC endocrine therapy [9, 16]. By IHC, about 75-80% of invasive BrC are positive for ER expression and the proportion of positive cells ranges from 1% to 100% [9, 10].

ER positive tumors are associated with well-differentiated, less aggressive tumors and, consequently, with a better outcome when compared with ER negative tumors. However, ER expression alone has a limited prognostic value [16].

Moreover, the ER expression is a strong predictive marker of response to endocrine therapies such selective ER modulators (e.g. tamoxifen) or aromatase inhibitors (e.g. letrozole) [9, 16].

Progesterone Receptor

PR expression suggests an active ER signaling pathway. The activation of PR by progesterone hormone stimulates cell growth. About 60-75% of invasive BrC are positive for PR by IHC and tumors PR positive are rarely ER negative [9, 16]. RP positive tumors

also associate with response to endocrine therapies. Moreover, tumors positive for ER but negative for PR are less responsive to these therapies than tumors positive for both hormone receptors [9, 16].

Erb-b2 receptor tyrosine kinase 2

Erb-b2 receptor tyrosine kinase 2 (ERBB2) gene [also known as *Human epidermal growth factor 2 receptor* (HER2)] is localized in chromosome 17 and codes for a growth factor receptor present in breast epithelial cell surface. This oncogene is amplified in 13% to 20% of BrC tumors with overexpression of protein. Moreover, about 55% of these cases do not express hormone receptors (ER and PR) [16]. The prognostic value of ERBB2 overexpression is complex once ERBB2 expression is associated with poor prognosis and at the same time to a favorable response to ERBB2-target therapies such as trastuzumab (a therapy targeting the ERBB2 protein) [9, 16].

Emerging prognostic assays

In the last decade, due to the limitations of current prognostic biomarkers, new prognostic tests have been developed. Multigene signatures were proposed as being useful complementary tools to the current biomarkers that may help in therapeutic decisions. These tests are, indeed, multivariate models to predict recurrences [17].

The first-generation signatures such as Oncotype DX, MammaPrint or Genomic Grade Index) are more accurate in predicting recurrence within the first 5 years' than in later years. Differently newer tests, such as PAM50, EndoPredict or Breast Cancer Index, hold a better prognostic value for late recurrence keeping the prognostic value for early recurrence [17].

However, the use of these signatures is limited and the evidence of clinical usefulness of some signatures is still weak. The American Society of Clinical Oncology (ASCO) already endorsed some of those signatures only for ER/PR-positive, ERBB2-negative and lymph node negative BrC. The ASCO 2016 guidelines only recognized the usefulness of OncotypeDX, EndoPredict, Breast Cancer Index and PAM50 to guide decision on adjuvant systemic therapy [18].

Treatment

The BrC treatment is a complex work-up which combines the TNM staging and the prognostic biomarkers status. The treatment decision in BrC should always be provided by a multidisciplinary clinical team and may include one or more therapeutic strategies, including surgery, radiotherapy, endocrine therapy, chemotherapy and target therapy [11].

Surgery

Surgery is usually the primary treatment for early-stage BrC. About 60-80% of BrC patients are amenable for conservative breast surgery (wide local excision) followed by radiotherapy [11, 19]. However, for some cases mastectomy remains as the best option due to: tumor size (relative to breast size), tumor multi-centricity, inability to achieve negative surgical margins after multiple resections, contraindications to radiotherapy, or simply for patient choice [11].

Radiation Therapy

Radiation Therapy is highly recommended after a conservative breast surgery. A specific planning for each patient is of major importance in order reduce the irradiation of adjacent organs and ensure the target dose delivery. Thus, it is recommend a computer tomography-based treatment planning [19].

The radiation of entire breast reduces the risk of recurrence by 15% in first 10 years and reduces the risk of mortality by BrC in 4% [11, 19]. In patients that undergo mastectomy and display positive lymph nodes, radiation therapy reduces the 10-year risk of any recurrence (including locoregional and distant) by 10% and the 20-year risk of breast cancer-related mortality by 8% [11].

Chemotherapy

Chemotherapy decision for BrC treatment is complex and takes into account the predicted sensitivity to particular treatment methods and benefit from its use along with individual risk of relapse [20]. Chemotherapy can be used as adjuvant therapy or as neoadjuvant therapy. Neoadjuvant chemotherapy is recommended in locally advanced BrC or in and large ‘operable’ cancers, in particular when mastectomy is required due to tumor size, since it may reduce the extent of surgery needed [11].

The benefit of adjuvant chemotherapy is higher in ER-negative BrC (either positive or negative for ERBB2). However, ER-positive BrC with ERBB2 amplification or high-risk of recurrence may also benefit of adjuvant chemotherapy [20, 21].

Endocrine Therapy

Endocrine Therapy is recommend to all BrC positive for hormone receptors. The prescription’s choice is primarily determined by patient’s menopausal status. In premenopausal, tamoxifen 20 mg/day for 5–10 years is standard. When patients become postmenopausal during the first 5 years of tamoxifen, a switch to an aromatase inhibitor (such as letrozole), seems to be particularly beneficial [11, 20]. High risk premenopausal patients should be considered for ovarian function suppression as well [21]. In postmenopausal patients, tamoxifen and aromatase inhibitors are valid options. Aromatase

inhibitors associate with increased disease-free survival, but no significant impact was observed in overall survival for treatment up to 5 years. Moreover, there is no proven benefit for the routine use of aromatase inhibitors for more than 5 years [11, 20].

Target Therapy

ERBB2 target therapy is the most used target therapy in BrC treatment in patients with ERBB2 overexpression/amplification. The monoclonal antibody against ERBB2, Trastuzumab, combined with chemotherapy reduces the recurrence risk by approximately 50% when compared with chemotherapy alone, translating into a 10% absolute improvement in long-term disease-free survival and 9% increase in 10- year overall survival [11, 20]. Trastuzumab is approved for treating patients with node-positive disease and for node-negative patients with tumors greater than 1 cm. However, due to high risk for recurrence, N0 staged patients with tumors less than 1 cm, should be also be considered for this treatment, particularly if they are ER-negative disease [11, 19].

MOLECULAR BIOLOGY OF BREAST CANCER

BrC is a multifactorial disease resulting from several genetics and epigenetics aberrations. In past decades, multiple efforts have been made in order to better understand BrC disease and improve the diagnostic, prognostic and predictive biomarkers and unsure a better treatment.

GENETICS

Several genetic aberrations are described in BrC, including point mutations, deletions, amplifications, rearrangements, translocations, and duplications. These DNA alterations result in oncogenes' activation and of tumor suppressors' silencing (Figure 6) [22].

PI3K and Ras-MEK signaling pathways are active by ERBB2 amplification in about 15% of BrC. The activation of these pathways leads to tumor cell growth and proliferation [10, 23]. Likewise, cell cycle control pathways are deregulated in BrC. In about 10-12% of the patients, *cyclin D1* is amplified which acts upon the Rb protein releasing E2F transcription factor that will induce cell cycle entrance [10, 23]. Moreover, DNA damage and genomic instability are associated with BrC. *BRCA1* and *BRCA2* mutations are associated with DNA repair deficiency and *TP53* mutations are associated with cell cycle checkpoints failure [10, 12, 23]. Single Nucleotide Polymorphisms (SNPs) and copy number variations (CNVs) can also have an impact in BrC incidence. Additionally to inherited mutations, SNPs and CNVs provide genetic background for BrC development [23].

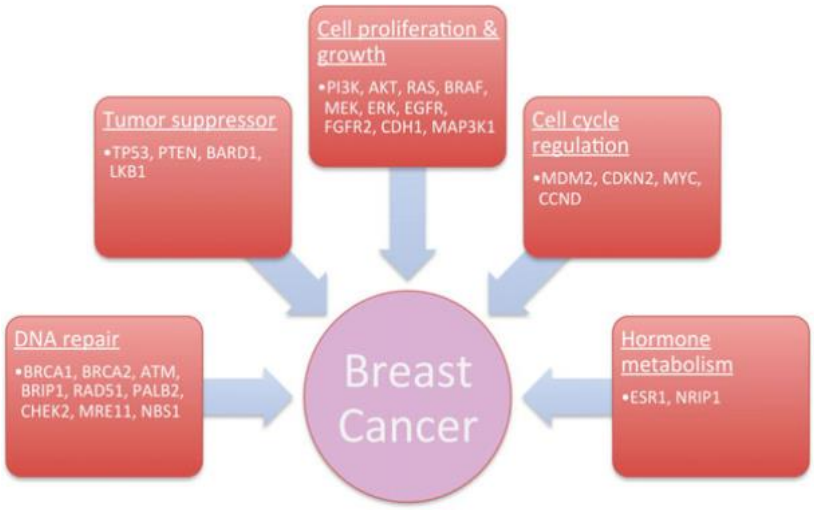

Figure 6 – Scheme of the mechanism related with BrC carcinogenesis with the most frequent genes involved. From [22]

Molecular Subtypes of Breast Cancer

BrC is a complex and heterogeneous disease in which tumors with the same histological types and identical stage often present divergent outcomes and therapeutic responses. In order to better understand these differences and due to the advances in high throughput technologies, multiples studies were made to adequately profile and classify BrC tumors [24, 25].

Microarray-based gene expression profiling endorsed specific molecular subtypes of BrC with coherent expression patterns, reflecting variation in the patient's prognosis [24, 26]. The molecular classification of BrC led to identification of widely accepted major subtypes that carry prognostic value (Figure 7)[26].

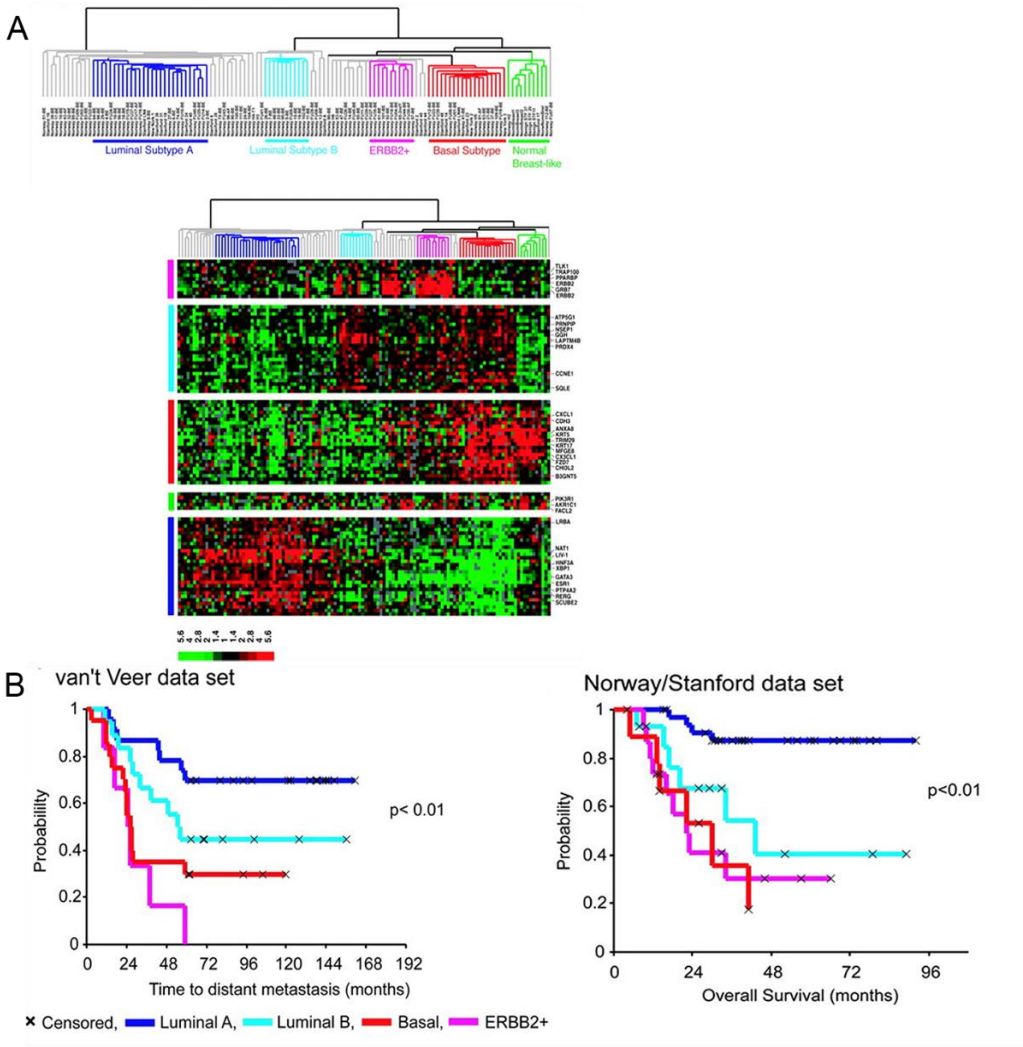

Figure 7 – Molecular intrinsic subtypes of Breast Cancer (A) Hierarchical clustering of 115 tumor tissues and 7 nonmalignant tissues using the "intrinsic" gene set. Experimental dendrogram showing the clustering of the tumors into five subgroups. Branches corresponding to tumors with low correlation to any subtype are shown in gray. (top panel) Gene cluster associated with each subtype (bottom panel).(B) Kaplan–Meier analysis of disease outcome in two patient cohorts. Time to development of distant metastasis in the 97 sporadic cases from van't Veer et al. (left panel). Overall survival for 72 patients with locally advanced breast cancer in the Norway cohort (right panel). Adapted from [26]

The first main division refers to the expression of luminal epithelial cells, including the ER and those that were negative for these genes. The ER positive/luminal subtype represents the majority of BrC (70-80%) and are generally associated with good prognosis [24, 26, 27]. Furthermore, luminal carcinomas can be subdivided in Luminal A and Luminal B subtypes according to ERBB2 expression and proliferation rates. Luminal B is associated with ERBB2 expression and/or high proliferation rates and, consequently, with worse prognosis than Luminal A tumors [26, 27].

In the non-luminal BrC, the ERBB2 overexpression subtype characterizes by high expression of several genes in the ERBB2 amplicon at 17q22.24 including *ERBB2* and *GRB7* [25]. These cancers usually display a poor prognosis due to high proliferation and poor differentiation, however can be a target of targeted therapies [25, 26].

Basal-like subtype is also a non-luminal BrC containing a heterogeneous group of tumors that may represent up to 15% of all BrC. Moreover, is characterized by strong expression of basal epithelial genes and low expression of luminal epithelial. Additionally, this subtype is associated with *BRCA1* mutations and to the lack of ER, PR and ERBB2 expression [24-27]. Moreover, over the years basal-like tumors became known as Triple Negative BrC (TNBC), although not all basal-tike BrC are TNBC and vice-versa. Seventy to eighty percent of TNBC correspond to all basal-like tumors, however the remaining 20-30% cases may be one of others subtypes [28]. Furthermore, other subtypes as normal-like and claudin-low carcinomas are also described as non-luminal carcinomas. Normal-like present similar gene expression pattern as the normal breast, and it is yet not clear whether represents a distinct subtype or is due to a technical artifact introduced by low tumor cell composition of the sampled specimen [26, 27]. Claudin-low carcinomas are usually TNBC characterized by the low to absent expression of luminal differentiation markers, high enrichment for epithelial-to-mesenchymal transition markers, immune response genes and cancer stem cell- like features [29].

Currently the assessment of the four main molecular intrinsic subtypes is being tested using IHC (Table 3) [11]. Over the years, studies assessing correlation between IHC findings, BrC patients' prognosis and gene expression profiling, specially, regarding the differentiation between luminal A and luminal B BrC have performed. Presently, TNBC are classified as basal-like carcinomas, however the search for new markers to a better classification of TNBC is currently undergoing [21, 30, 31].

Table 3 – Surrogate definitions of intrinsic subtypes of Breast Cancer according to European Society for Medical Oncology (ESMO) Clinical Practice Guidelines for diagnosis, treatment and follow-up to primary Breast Cancer. Adapted from [11].

Intrinsic subtype	Clinicopathologic surrogate definition	Notes
Luminal A*	'Luminal A-like' • ER-positive • ERBB2-negative • Ki67 low** • PR high**	*if molecular signature is available, a low-risk signature is associated with Luminal A BrC and high-risk signature with Luminal B BrC **Scores should be interpreted in the light of local laboratory values ***There is ~80% overlap between 'triple-negative' and intrinsic 'basal-like' subtype, but 'triple-negative' also includes some special histological types such as (typical) medullary and adenoid cystic carcinoma with low risks of distant recurrence.
Luminal B*	'Luminal B-like (ERBB2-negative)' • ER-positive • ERBB2-negative • And either Ki67 high or PR low 'Luminal B-like (ERBB2-positive)' • ER-positive • ERBB2-positive • Any Ki67 and any PR	
ERBB2 overexpression	'ERBB2-positive (non-luminal)' • ERBB2-positive • ER and PR absent	
Basal-like	'Triple-negative (ductal)'*** • ER and PR absent • ERBB2-negative	

ER – Estrogen Receptor; ERBB2 – Erb-b2 receptor tyrosine kinase 2; PR – Progesterone Receptor; BrC – Breast Cancer

EPIGENETICS

In 1942, Conrad Waddington defined epigenetics as a biology branch which studies the interaction between genes and their products and bring the phenotype. However, the concept of epigenetics has been changed since then due to increased knowledge on this field of research. Epigenetics is, nowadays, recognized as the field of study of heritable chances in gene expression that occurs without alteration in DNA sequence. In 2010, the NIH Road-map Epigenomics Project defined epigenetics as "heritable changes in gene activity and expression (in the progeny of cells or of individuals) and also stable, long-term alterations in the transcriptional potential of a cell that are not necessarily heritable"[32].

Epigenetic Mechanisms

Epigenetic mechanisms can be grouped into four main types: DNA methylation, non-coding RNAs, histone post-translational modifications and histone variants (Figure 8). These mechanisms are essential for the normal function of organism. The disruption of one

or more of epigenetic mechanism can result in abnormal gene expression, altering the cell homeostasis which conducts to disease development including cancer [32, 33].

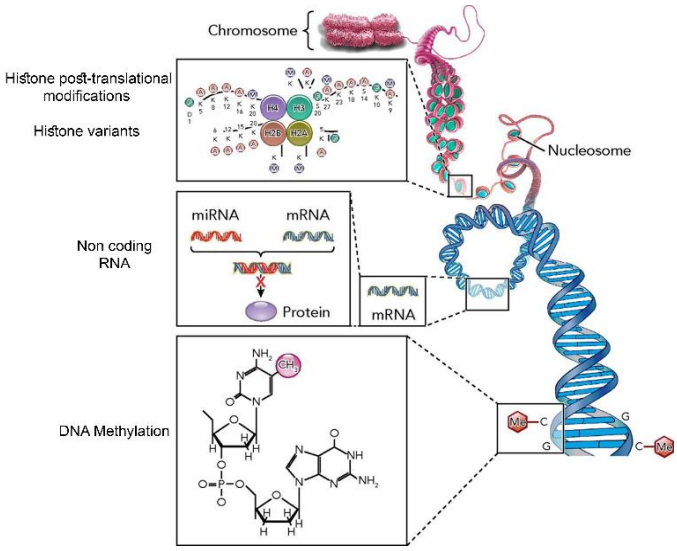

Figure 8 – Scheme of four main mechanisms of epigenetic regulation. Adapted from [33]

Non-coding RNAs

Non-coding RNAs (ncRNAs) are a class of RNA that do not encode a protein but are considered active participants in gene expression regulation. The ncRNAs can be further divided according their size into large ncRNAs (more 200 nucleotides) and small ncRNAs (less 200 nucleotides) [34]. The most widely known ncRNAs are the microRNA (miRNAs). MiRNAs have a 17-25 nucleotides involved in several biological processes such as cellular development, differentiation, apoptosis, proliferation, tumor growth among others [32]. Aberrant expression of miRNAs was associated with multiple cancers, including BrC [35].

Histone Post-translational Modifications and Variants

Eukaryotic chromosomes are composed of essential units of organization, the nucleosomes which compose the chromatin. Each nucleosome is composed of an octamer of four core histones (H2A, H2B, H3, H4) that is wrapped by a 147-bp stretch of DNA [35]. Serving as the basic modules for DNA packaging within a cell, nucleosomes regulate gene expression by altering the accessibility of regulatory DNA sequences to transcription factors. The incorporation of histone variants, e.g. H3.3 and H2A.Z, into nucleosomes influences nucleosome occupancy and thus in gene activity [36].

Additionally, histone can suffer post-translation modifications which occur mainly in the terminal tails. The most studied post-translational modifications are methylation, acetylation and phosphorylation. Those post-translational modifications regulate essential cellular process such as replication, transcription and repair by either changing the chromatin's availability or by recruiting and/or occluding non-histone effector proteins, which

decode the message encoded by the modification patterns [34, 36]. Importantly, both, histone variants and histone post-translational modifications have been implicated in tumorigenesis [34, 35].

DNA Methylation

DNA methylation is one the most frequent and well-studied epigenetic events.

The DNA methylation consists in a covalent chemical modification resulting in addition of a methyl (CH_3) group at the carbon 5 position of the cytosine ring of CpG dinucleotides (cytosines that precede guanines). This reaction is carried out by DNA methyltransferases (DNMTs) using S-adenosyl methionine as a methyl donor [34].

In mammalian cells there are five described DNMTs, however only three seems to possess methyltransferase activity: DNMT1, DNMT3a and DNMT3b [37]. DNMT1 is classified as *maintenance* DNMT once it is responsible of recognizing hemimethylated CpG sites and to ensure the proper inheritance of DNA methylation patterns from mother to daughter strand [37, 38]. DNMT3a and DNMT3b are *de novo* DNMTs and responsible for the establishment of the tissue-specific DNA methylation patterns following zygote implantation. In normal conditions these enzymes are highly expressed in undifferentiated embryonic stem cells but are down-regulated after differentiation [34, 38].

About 70-80% of CpG sites are methylated, their distribution in human genome is not random and their methylation probabilities depend highly on their position within the genome [32, 38]. Genomic regions with high CG content and high frequency of CpG dinucleotides are defined as CpG islands [38]. However, DNA methylation does not occur exclusively at CpG islands. Indeed, it may also occur in CpG island shores, which are regions of lower CpG density that lie in close proximity (~2 kb) of CpG islands [37].

In normal cells, active promoters and CpG islands are largely resistant to DNA methylation. In opposite, repetitive DNA elements, pericentromeric regions of the genome and gene bodies, are often heavily methylated. Moreover, CpG island shores' methylation is also associated with transcriptional inactivation. Most of the tissue-specific DNA methylation seems to occur at CpG island shores (Figure 9, left panel) [37, 38].

In cancer cells global hypomethylation is accompanied by hypermethylation of localized promoter associated CpG islands that usually remain unmethylated in normal cells (Figure 9 right panel). Global DNA hypomethylation plays a significant role in tumorigenesis and occurs at various genomic sequences usually methylated, such as repetitive DNA elements, retrotransposons and gene bodies. DNA hypomethylation is associated to increased genomic instability by promoting chromosomal rearrangements. Hypomethylation at specific promoters can activate the aberrant expression of oncogenes promoting tumor growth and proliferation [34, 36, 37]. Simultaneously, DNA

hypermethylation typically affects promoter CpG islands at specific genes inactivating their transcription [34, 37]. Indeed, hypermethylation usually affects tumor suppressor genes involved in the main cellular pathways such as DNA repair, vitamin response, Ras signaling, cell cycle control and apoptosis, among others [34, 37].

Hypermethylated promoters have been proposed as a new generation of biomarkers since they hold great diagnostic and prognostic promise for clinicians [34].

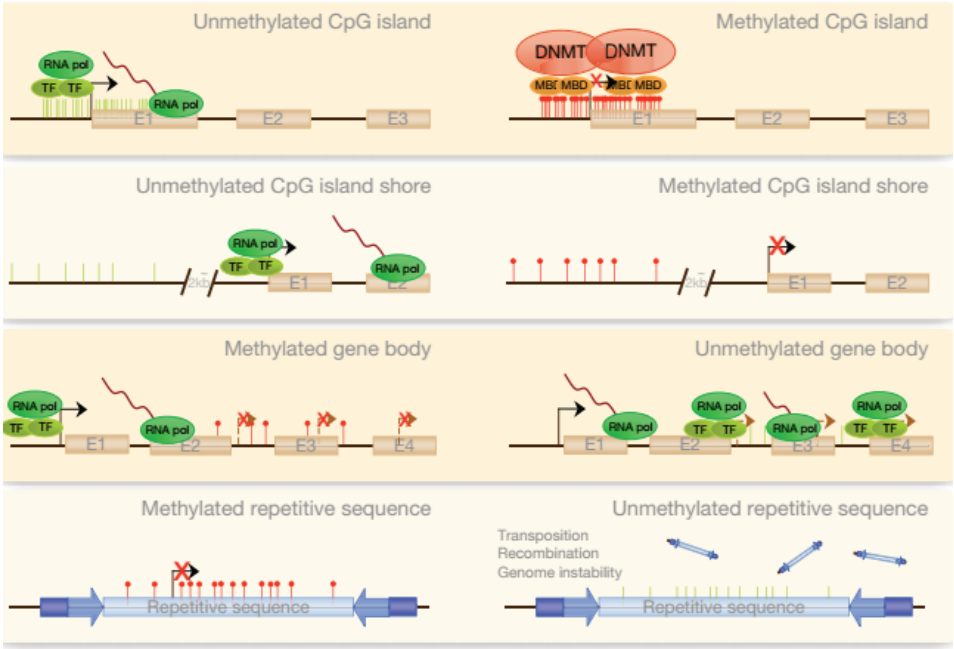

Figure 9 – DNA methylation patterns. DNA methylation can occur in different regions of the genome. The left panel show the normal scenario in mammalian cells. The right panel show the alteration common in cancer cells. Adapted from [37]

DNA Methylation and Breast Cancer

DNA methylation patterns may differ largely between tumor and normal cells and numerous studies reported its important role in carcinogenesis. Similarly to other malignancies, BrC globally display a hypomethylation pattern associated with lead to oncogenes' expression. Nonetheless, promoter hypermethylation of several tumor suppressor genes are more commonly described in BrC [39]. In fact, DNA global hypomethylation and tumor suppressor gene promoter hypermethylation have been reported occurring simultaneously as independent processes in breast carcinogenesis and at different stages of BrC, being promoter hypermethylation described as an early event in BrC directly linked to tumor development [39].

The knowledge of DNA methylation implication in tumor initiation and progression in BrC unveiled the potential of this epigenetic alterations as diagnostic, prognostic and predictive biomarkers [40, 41].

From the several genes reported, the below mentioned were already proposed as putative BrC biomarkers.

Adenomatosis polyposis coli (APC)

APC is a tumor suppressor gene that is involved in cell signaling through Wnt pathway and plays an important role in cell-cycle regulation and apoptosis [42, 43]. High levels of methylation in *APC* promoter has been described in samples from several neoplasms such breast, colon, prostate and lung. In BrC, *APC* promoter hypermethylation seems to be associated with menopausal status and is more frequent in tumors ER and/or ERBB2 positive [42-45].

BRCA1, DNA repair associated (BRCA1)

BRCA1 is a tumor suppressor gene important for DNA repair, control of chromatin remodeling and transcription, and cell cycle regulation. The *BRCA1* promoter methylation is a rather uncommon event in BrC [46], however, is was associated with basal-like/TNBC subtype [47, 48]. Moreover, methylation might also be detected in serum of BrC patients but not in healthy patients [49].

Cyclin D2 (CCND2)

CCND2 is a tumor suppressor gene which is involved in cell cycle regulation. This gene was found hypermethylated in BrC samples and may constitute a good diagnosis biomarker [45, 50]. The *CCND2* promoter methylation has been more frequently found in ER positive tumors [43], which represent the majority of breast cancer lesions.

Fork-head box A1 (FOXA1)

FOXA1 is a member of the FOX family of transcription factors that is comprised at least 40 members. It is a key forkhead transcription factor that can directly bind condensed chromatin, displace repressive linker histones and recruit other transcription factors to promote transcription [51, 52]. ERα recruits *FOXA1* as a cofactor for chromatin binding and has been suggested to be responsible for the regulation of nearly 50% ER-target genes [52]. In BrC, the hypermethylation of *FOXA1* promoter has been associated with basal-like carcinoma [52, 53].

Ras association domain family 1 isoform A (RASSF1A)

RASSF1A is a tumor suppressor gene frequently silenced by respective promoter hypermethylation in several human cancers. This gene is implicated in multiple pathways that control cell-cycle, apoptosis and cell motility and invasion [54]. Similarly to *APC*, *RASSF1A* promoter was reported to be frequently hypermethylated in several neoplasms. In BrC, *RASSF1A* promoter hypermethylation was associated with menopausal status and with early stages of BrC development [42, 50, 55]. Similarly, *RASSF1A* inactivation by

methylation is more frequently found in ER and/or ERBB2 positive tumors and was reported to have predict overall survival and disease-free survival [43, 50, 56, 57].

Secretoglobin family 3A member 1 (SCGB3A1/HIN1)

SCGB3A1, also known as HIN1, is described as a tumor suppressor gene involved in cell cycle control, migration and invasion suppression and apoptosis stimulation [58]. Moreover, higher *SCGB3A1* methylation levels were associated with ER positive samples and well-differentiated tumors [59, 60], whereas low methylation levels were associated with basal-like subtype [45, 48, 59].

AIMS

AIMS

BrC is a heterogeneous disease involving genetic and epigenetic modifications. Despite the technical advances for early detection, BrC still is the main cause of cancer death among women in Europe. More efficient detection tests, able to diagnose and characterize BrC may improve the treatment efficacy in BrC patients.

Epigenetic modifications occur at the onset, development and progression of BrC. DNA methylation markers have, thus, emerged as potential diagnostic and prognostic biomarkers due to their stability and easy assessment by Polymerase Chain Reaction (PCR) technologies. Thus, the major aim of this dissertation is to develop a DNA methylation-based test capable of early detecting all the major BrC molecular subtypes.

Thus, the specific aims of this work were:

- Assess the methylation status of six genes' promoters (*APC, BRCA1, CCND2, FOXA1, RASSF1A* and *SCGB3A1*) in a series of BrC and normal breast (NBr) frozen tissues.
- Define the BrC molecular subtype using IHC and correlate it with the methylation levels of the previously indicated six genes' promoters.
- Evaluate the association of methylation levels with standard clinicopathological parameters and assess the prognostic value of aberrant methylation of those six genes' promoters in a cohort of BrC patients.

MATERIAL AND METHODS

PRELIMINARY DATA

IHC was performed to identify the molecular subtype of each case BrC included in this study. Commercially available antibodies were used for ER (Clone 6F11, mouse, Leica), PR (Clone 16, mouse, Leica), ERBB2 (Clone 4B5, rabbit, Roche) and Ki67 (Clone MIB-1, mouse, Dako).

IHC was carried out in BenchMark ULTRA (Ventana, Roche) using ultraView Universal DAB Detection Kit (Ventana, Roche) according with manufacturer's instructions.

Each case was evaluated by an experienced pathologist. IHC staining was classified according to College of American Pathologists recommendations. For ERBB2 immunostaining, the cases score 2+ were confirmed by fluorescence *in situ* hybridization. Each case was categorized according the guidelines of ESMO (Table 3). The cut-off for Ki67 was 15% of positive cells and the cut-off for high *vs.* low expression of PR was 25%.

PATIENTS AND SAMPLES COLLECTION

For this study, 137 BrC samples were prospectively collected from patients submitted to surgery as 1st treatment from 1996 to 2001, at the Portuguese Oncology Institute of Porto. All control samples (NBr) were collected from reduction mammoplasty of contralateral breast of BrC patients without breast cancer hereditary syndrome. After surgical resection and examination, samples were immediately frozen at -80°C. Relevant clinical and pathological data was retrieved from the patient's clinical charts.

Five µm frozen sections were cut and stained by hematoxilin-eosin staining and histological evaluation was performed by an experienced pathologist.

DNA EXTRACTION

DNA extraction was performed by phenol-chloroform method. Ten µm sections from fresh-frozen tissues were cut and placed in 15mL tubes. To the fresh-frozen sections 2.700 mL of SE buffer (75mM NaCl and 25 mM EDTA), 300 µL of 10% SDS and 25 µL of proteinase K [20mg/mL(NZYTECH, Portugal)] were added. Samples were incubated at 55ºC until complete digestion was achieved. Additional proteinase K was added at twice a day.

Subsequent to digestion, all samples were transferred to Phase Lock Light 15mL tubes (5 Prime, Germany) previously centrifuged at 4000 rpm for 5 minutes and mixed with 3mL of phenol-chloroform pH=8 (Sigma-Aldrich, USA).

After centrifugation, at 4000 rpm for 20 minutes, the aqueous phase was transferred to new 15mL tube. For DNA precipitation, 6mL of cold absolute ethanol (Merkmilipore, Germany) and 1mL of ammonia acetate at 7.5M (Sigma-Aldrich, USA) were added to samples. After mixing, samples were placed at -20°C overnight.

Following, samples were centrifuged at 4000 rpm for 20 min and washed in 6mL ethanol 70% twice. The air dried pellets and eluted in sterile distilled water (B.Braun, Melsungen, Germany). DNA concentration and purity were assessed using NanoDrop Lite Spectrophotometer (Nanodrop Technologies, USA) and stored at -20°C until further use.

METHYLATION ANALYSIS

Sodium Bisulfite Modification

Sodium Bisulfite modification is a crucial step for all the methylation studies involving PCR and sequencing[61]. This process allows to differentiate methylated from unmethylated cytosines. The unmethylated cytosines suffer a sulphonation, deamination and desuphonaltion processes and are convert to uracil while methylated cytosine remains unaltered.

Genomic DNA from all samples was modified by sodium bisulfite using EZ DNA Methylation-Gold™ Kit (Zymo Research, USA) according to the manufacturer's guidelines (Appendix III).

To sodium bisulfite modification, it was used 1µg of DNA obtained from fresh frozen sections. In the end of protocol modified DNA was eluted with 60µL of sterile distilled water and stored at -80°C until further use.

One µg of CpGenome™ Universal Methylated DNA (Millipore, USA) was also modified, according to the method described above and eluted in 20µL of M-elution buffer.

Quantitative Methylation-Specific PCR (qMSP)

To analyze methylation levels in our samples was performed quantitative real-time methylation specific PCR (qMSP). Methylation specific PCR (MSP) allows the specific amplification of methylated or unmethylated alleles[62]. In MSP, the modified DNA is used as template and amplify using primers that containing or not CpGs, however MSP is a qualitative method [62, 63]. The qMSP is a quantitative method that combines the MSP and real-time PCR principles. This method is more sensitive and uses primers with CpGs allows the specific amplification of methylated DNA.

The modified DNA was used as template and samples were submitted to qMSP reactions for the target genes: *APC, BRCA1, CCND2, FOXA1*, *RASSF1A* and *SCGB3A1*; and for *Actin* β (*ACTβ*) gene used as reference gene.

Reactions were performed in 96-well plates using Applied Biosystems 7500 Real-Time System (Thermo Fisher Scientific, USa). Briefly, per each well were added 2 µL of modified DNA and 5 µL of 2X KAPA SYBR FAST qPCR Master Mix. The primer volume used was 0.3µL from a 10mM solution of forward and reserve primers, except to *ACTβ*

which primer volume used was 0.4µL. Sterile distilled water was added until 10 µL of reaction volume were achieved. Primer sequences are listed in Table 4.

Table 4 – Primers sequences and qMSP conditions for each gene studied.

Gene	Primers	Annealing Temperature ºC
ACTβ	F – 5' TGG TGA TGG AGG AGG TTT AGT AAG T 3'	60ºC
	R – 5' ACC AAT AAA ACC TAC TCC TCC CTT AA 3'	
APC	F – 5' TGT GTT TTA TTG CGG AGT GC 3'	62ºC
	R – 5' CAC ATA TCG ATC ACG TAC GC 3'	
BRCA1	F – 5' TTT CGT GGT AAC GGA AAA GC 3'	60ºC
	R – 5' ATC TCA ACG AAC TCA CGC CG 3'	
CCND2	F – 5' TTT GAT TTA AGT ATG CGT TAG AGT ACG 3'	62ºC
	R – 5' ACT TTC TCC CTA AAA ACC GAC TAC G 3'	
FOXA1	F – 5' CGA CGT TAA GAC GTT TAA GC 3'	62ºC
	R – 5' CGC TCA ACG TAA ACA TCT TAC 3'	
RASSF1A	F – 5' AGC GAA GTA CGG GTT TAA TC 3'	60ºC
	R – 5' ACA CGC TCC AACC GA ATA 3'	
SCGB3A1	F – 5' GTA CGG TCG TGA GCG GAG C 3'	64ºC
	R – 5' GAA ACT TCT TAT ACC CGA TCC TC 3'	

The PCR program consisted of a period of 3 minutes at 95°C for enzyme activation followed by 45 cycles with 3 seconds at 95°C (for DNA denaturation) and 30 seconds at the temperature indicated in Table 4 (for annealing, extension and data acquisition).

All the samples were run in triplicated and in each plate 2 wells contained a negative control. The modified CpGenome™ Universal Methylated DNA was used as positive control and it was diluted in five serial dilutions by a 5x dilution factor. These serial dilutions were run in each plate and were used to generate a standard curve thus allowing absolute quantification as well as ascertaining PCR efficiency. All plates had an efficiency between 90-100%.

For each gene, relative methylation levels were calculated as the ratio between the target gene mean quantity and *ACTβ* mean quantity:

$$Relative\ Methylation\ levels = \frac{Mean\ quantity\ of\ target\ gene}{Mean\ quantity\ of\ ACT\beta} \times 1000$$

STATISTICAL ANALYSIS

The frequency, median and interquartile range of promoter methylation levels of normal tissue/control samples were determined.

Non-parametric tests were performed to determine statistical significance in all the comparisons made. In particular, Kruskall-Wallis test was used in comparisons between 3 or more groups, whereas Mann-Whitney U test was used for comparisons between two groups.

Logistic regression models were also built in order to evaluate the potential of using the targets as a panel to increase performance. To assessing the targets' performance as biomarkers, ROC curves were built. Moreover, biomarker parameters [specificity, sensitivity, positive predictive value (PPV), negative predictive value (NPV) and accuracy] were determined using the formulas provided in Table 5. For this, the cut-off established was the highest value obtained by the ROC curve analysis [sensitivity + (1-specificity)]. To categorize samples as methylated or unmethylated, a cutoff value was chosen based on cutoff obtained by the ROC curve analysis for each gene.

Spearman nonparametric correlation test was performed to correlate methylation levels and age.

Table 5 – Formulas used for biomarker parameters calculation.

Tumor vs. Control					
	Tumor	Control		**Sensitivity (%)**	(C/A) x100
Total	A	B		**Specificity (%)**	(F/B) x 100
> cut-off	C	D		**PPV (%)**	(C/(C+D)) x 100
< cut-off	E	F		**NPV (%)**	(F/(E+F)) x 100
				Accuracy (%)	[(C+F)/(A+B)] x 100

Disease-specific survival curves and disease-free survival curves (Kaplan–Meier with log rank test) were computed for standard clinicopathological variables and for categorized methylation status. A Cox-regression model comprising all significant variables (multivariable model) was computed to assess the relative contribution of each variable to the follow-up status.

Two-tailed P-values were derived from statistical tests, using a computer assisted program (SPSS Version 20.0, Chicago, IL), and results were considered statistically significant at $P < 0.05$, with Bonferroni´s correction for multiple tests, when applicable

Graphics were assembled using GraphPad 6 Prism (GraphPad Software, USA).

RESULTS

CLINICAL SAMPLES

For the purpose of this study, clinical tissue samples of NBr and BrC, stored at the Department of Pathology were used after institutional ethical committee's approval (CES 120/015). The relevant clinical and pathological data are presented in Table 6.

Table 6 – Clinical and pathological data of al Breast Cancer (BrC) and Normal Breast (NBr) samples used in this study.

Clinicopathologic features	BrC	NBr
Patients (n)	137	18
Age median (range)	62 (33-88)	54 (40-70)
Molecular subtype (no).		n.a.
Luminal A	29	
Luminal B	74	
ERBB2 overexpression	12	
TNBC	22	
Histological Type		n.a.
Invasive Ductal Carcinoma	116	
Invasive Lobular Carcinoma	8	
Special Subtype Carcinomas	5	
Mixed Type Carcinoma	8	
Grade (no.)		n.a.
G1	13	
G2	63	
G3	56	
Not determined	5	
Estrogen Receptor Status		n.a.
Positive	35	
Negative	102	
Progesterone Receptor Status		n.a.
Positive	66	
Negative	71	
ERBB2 Receptor Status		n.a.
Positive	25	
Negative	112	
Pathological T Stage (no.)		n.a.
pT1	42	
pT2	84	
pT3	7	
pT4	2	
pTx	2	
Pathological N Stage (no.)		n.a.
pN0	52	
pN1	45	
pN2	17	
pN3	19	
pNx	4	

Stage (no.)		n.a.
I	21	
II	71	
III & IV	41	
Not determined	4	

BrC – Breast Cancer; NBr – Normal Breast; n.a. – non applicable; TNBC – Triple Negative Breast Cancer;

ASSESSMENT OF THE METHYLATION LEVELS IN BRC AND NBR TISSUE SAMPLES

To verify whether the methylation levels of selected genes were cancer-specific, it was evaluated in BrC and NBr tissue samples. The majority of BrC samples presented higher levels of methylation at *APC*, *CCND2, FOXA1, RASSF1A* and *SCGB3A1* promoters when compared to NrB samples ($p<0.0001$, $p<0.0001$, $p=0.0016$, $p<0.0001$ and $p<0.0001$, respectively) (Table 7), whereas for *BRCA1*, no differences were depicted, and, thus, *BRCA1* was not further tested as diagnostic biomarker.

Table 7 – Frequency of positive cases [n(%)] and distribution of methylation levels of cancer-related genes [gene/ACTB x1000 median (IQR[1])]

Genes	NBr		BrC		*p* value
	n (%)	Median (IQR)	n (%)	Median (IQR)	
APC	0/18 (0%)	1.053 (0.548-2.049)	70/137 (51%)	24.010 (1.423-232.219)	**<0.0001**
BRCA1	15/18 (83.3%)	0.116 (0.056-0.297)	135/137 (98.5%)	0.137 (0.052-0.395)	0.7627
CCND2	1/18 (5.6%)	0.064 (0-0.325)	93/137 (67.9%)	5.551 (0.345-31.217)	**<0.0001**
FOXA1	0/18 (0%)	13.595 (8.822-21.568)	65/137 (47.4%)	28.325 (13.595-59.779)	**0.0016**
RASSF1A	1/18 (5.6%)	12.653 (4.056-21.265)	113/137 (82.5%)	329.841 (144.128-559.82)	**<0.0001**
SCGB3A1	0/18 (0%)	0.625 (0.264-3.078)	95/137 (69.3%)	217.267 (3.327-575.690)	**<0.0001**

NBr – Normal Breast; BrC – Breast Cancer; IQR – Interquartile Range

EVALUATION OF THE BIOMARKER DIAGNOSTIC PERFORMANCE

Because the majority of the tested genes showed higher methylation levels in BrC than in NBr samples (Figure 10, left panel), ROC curves (Figure 10, right panel) were constructed and empirical cut-off value was determined for each gene (Appendix IV).

[1] IQR – Interquartile Range

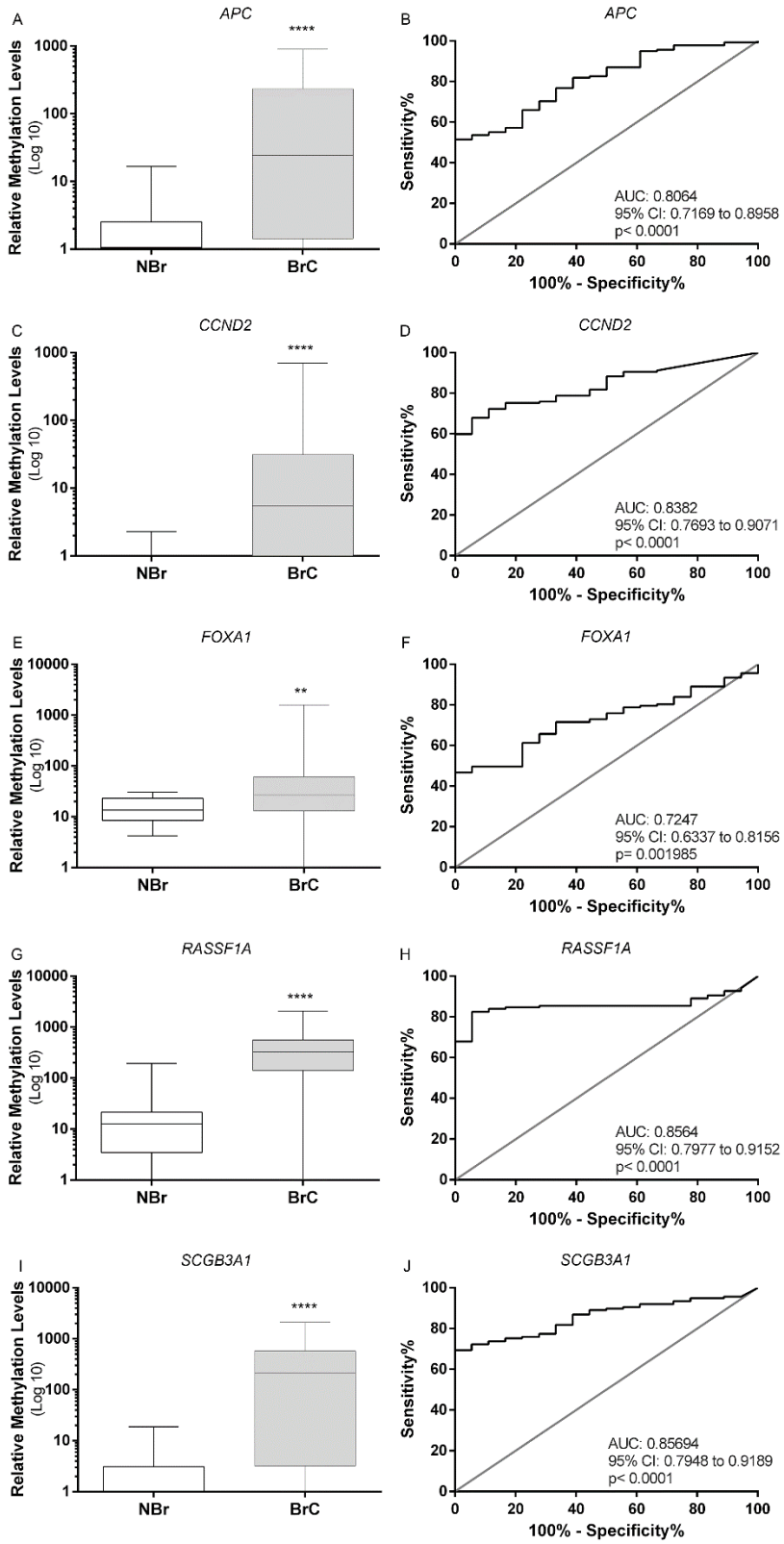

Figure 10 – Box-plots (left panels) of methylation levels in Normal Breast (NBr) and Breast Cancer (BrC) samples of *APC* (A), *CCND2* (C), *FOXA1* (E), *RASSF1A* (G) and *SCGB3A1* (I) and respective Receiver Operating Characteristic Curve (right panel) (B, D, F, H and J). Abbreviations: AUC – Area Under the Curve

Notably, except for *FOXA1*, all the genes displayed an Area Under the Curve (AUC) higher than 0.80. The empirical cut-off for each gene allowed for the calculation of performance for each gene individually (Table 8). *APC*, *FOXA1* and *SCGB3A1* showed 100% specificity for cancer detection, however the negative predictive values (NPV) were not higher than 30%. RASSF1A presented the highest sensitivity (82.48 %) and the highest accuracy (83.87%).

Table 8 – Performance of promoter gene methylation as biomarkers for detection of Breast Cancer in tissue samples

Genes	Sensitivity %	Specificity %	PPV %	NPV %	Accuracy %
APC	51.09	100.00	100.00	21.18	56.77
CCND2	67.88	94.44	98.94	27.87	70.97
FOXA1	47.45	100.00	100.00	20.00	53.55
RASSF1A	82.48	94.44	99.12	41.46	83.87
SCGB3A1	69.34	100.00	100.00	30.00	72.90

PPV – Positive Predictive Value; NPV – Negative Predictive Value

The combination of multiple genes was tested, as well, in an attempt to improve the performance. The best results were achieved by the panel comprising *APC, FOXA1* and *RASSF1A* which showed an accuracy of 94.19% and an AUC of 0.9570 (Table 9, Figure 11). This three-gene panel detected BrC with 94.16% sensitivity, 94.4% specificity and a remarkable PPV.

Table 9 – Performance of promoter gene methylation as biomarkers for detection of Breast Cancer in tissue samples

Genes	Sensitivity %	Specificity %	PPV %	NPV %	Accuracy %
APC/FOXA1/RASSF1A	94.16	94.44	99.23	68.00	94.19

PPV – Positive Predictive Value; NPV – Negative Predictive Value

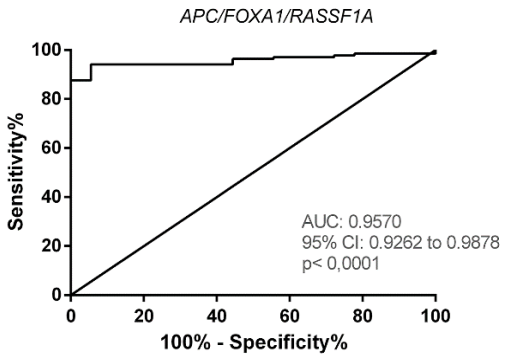

Figure 11 – Receiver Operating Characteristic Curve of the 3-gene panel (*APC, FOXA1* and *RASSF1A*) in Breast Cancer.

EXPLORING EPIGENETIC BIOMARKERS AS MOLECULAR SUBTYPE DISCRIMINATORS

Aiming to discriminate the four major molecular subtypes of BrC, the distribution of methylation levels for the six genes according to molecular subtype was carried out (Figure 12).

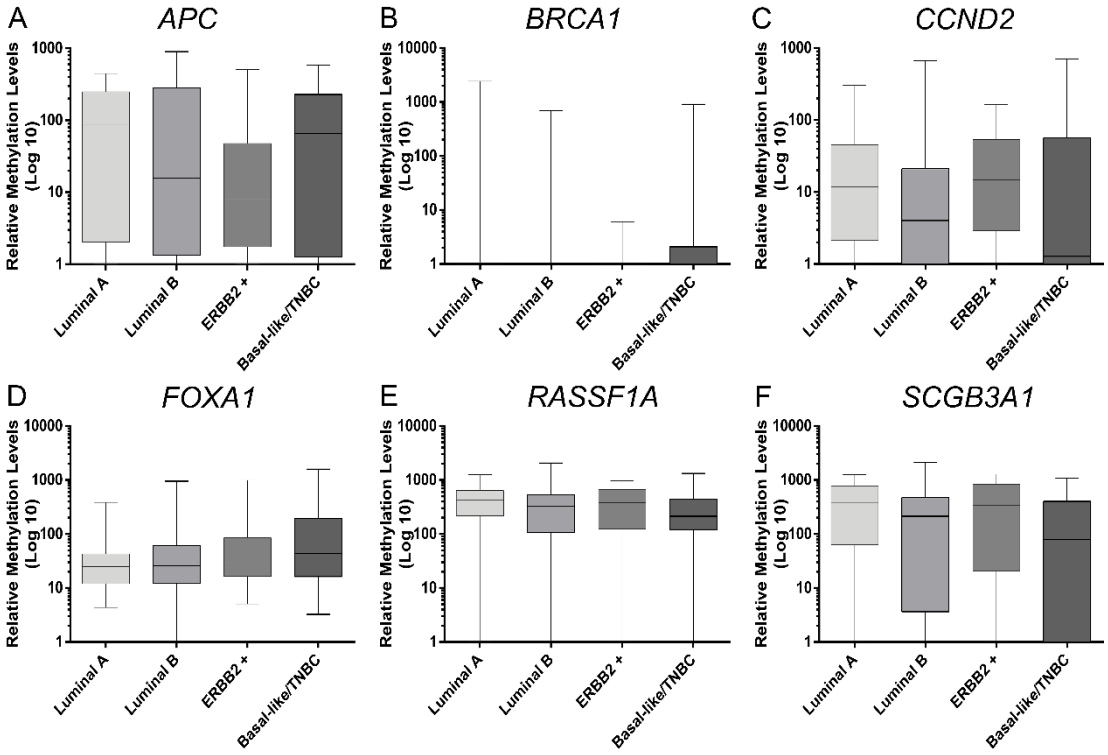

Figure 12 – Boxplots of *APC* (A), *BRCA1* (B), *CCND2* (C), *FOXA1* (D), *RASSF1A* (E) and *SCGB3A1* (F) methylation levels in the Breast Cancer molecular subtypes.

However, no significant differences in the promoter methylation levels of the target genes were apparent among the molecular subtypes.

ASSOCIATION BETWEEN PROMOTER METHYLATION LEVELS AND STANDARD CLINICOPATHOLOGICAL PARAMETERS

APC, *BRCA1, CCND2, FOXA1, RASSF1A* and *SCGB3A1* promoter methylation levels did not associate with tumor grade, pathological stage or ERBB2 status. Nevertheless, a significant correlation was found between *CCND2* and *RASSF1A* methylation levels and BrC patients' age (R=0.194, *p*=0.023 and R=0.223, *p*=0.009, respectively). This correlation was not apparent in controls. Additionally, a significant association was found between histological subtypes and *APC* and *SCGB3A1* methylation levels (Figure 13). The group of Special Subtype Carcinomas (SSC) showed the lowest

SCGB3A1 methylation levels in comparison to all the other histological types (Figure 13B) (p= 0.0126, p= 0.0446 and p= 0.0225, respectively for IDC, ILC and MTC). Moreover, SSC showed lower *APC* methylation levels than all the others histological types (Figure 13A), but significant differences were only attained for Invasive Lobular Carcinomas (ILC), (p=0.0293) (Figure 13B).

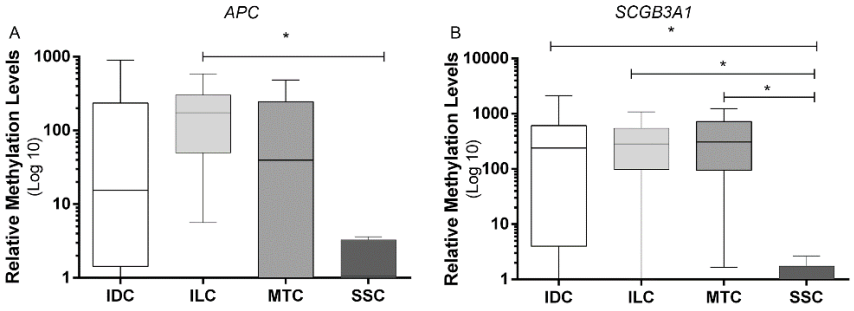

Figure 13 – Boxplots of *APC* (A) and *SCGB3A1* (B) methylation levels across the different Breast Cancer histological types. Abbreviations: IDC – Invasive Ductal Carcinoma; ILC – Invasive Lobular Carcinoma; MTC – Mixed Type Carcinoma; SSC – Special Subtype Carcinoma

FOXA1 and *RASSF1A* methylation levels associated with Hormone Receptor Status (Figure 14). BrC positive for both hormone receptors (ER and PR) displayed significantly lower FOXA1 methylation levels than ER and PR negative BrC samples (p= 0.0084). The same was observed for BrC samples that were only ER positive (p= 0.0319) (Figure 14 A). Contrarily, BrC samples positive for both Hormone Receptors showed higher *RASSF1A* methylation levels than ER positive only BrC samples (p= 0.0017). However, no statistical differences were observed between hormone receptor positive samples and hormone receptor negative samples (Figure 14 B). For the remaining genes no associations were disclosed regarding hormone receptor status.

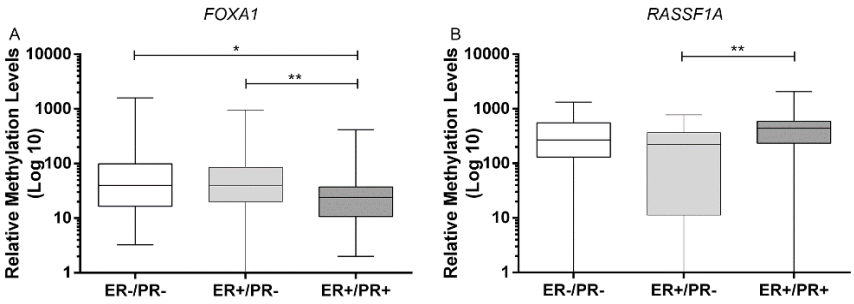

Figure 14 – Boxplots of the methylation levels of *FOXA1* (A) and *RASSF1A* (B) regarding Hormone Receptor Status. Abbreviations: ER+ – Estrogen Receptor Positive; ER- – Estrogen Receptor Negative; PR+ – Progesterone Receptor Positive; PR- – Progesterone Receptor Negative

SURVIVAL ANALYSIS

Survival analysis was performed aiming to evaluate the potential prognostic value of the six genes' promoters (*APC*, *BRCA1, CCND2, FOXA1, RASSF1A* and *SCGB3A1*) methylation levels. The survival analysis was carried out at 10 years of follow-up. At the time of the last follow-up 92 patients (67.2%) were alive with no evidence of cancer, 8 patients (5.8%) were alive with cancer and 37 patients (27.0%) had deceased, 24 of which due to BrC (17.5% of all cases).

Due to the reduce number of events and/or cases in same categories, some clinicopathological features were grouped. Grade was grouped in two categories (G1&G2 vs. G3), pT stage was grouped in three categories (pT1, pT2 and pT3&pT4), pN Stage was grouped in two categories (N0&N1 vs. N2&N3) and stage was grouped in three categories (I, II and III&IV).

Although, methylation levels of any of the studied genes did not associate with disease-free survival, statistically significance was found for tumor grade (p=0.041), pN stage (p<0.001) and Stage (p=0.038).

Concerning disease-specific survival the only gene that presented prognostic value was *FOXA1* categorized by percentile 75 of promoter methylation levels (Figure 15 A). Thus, BrC patients with *FOXA1* high methylation levels showed shorter disease-specific survival compared to BrC patients with low methylation levels.

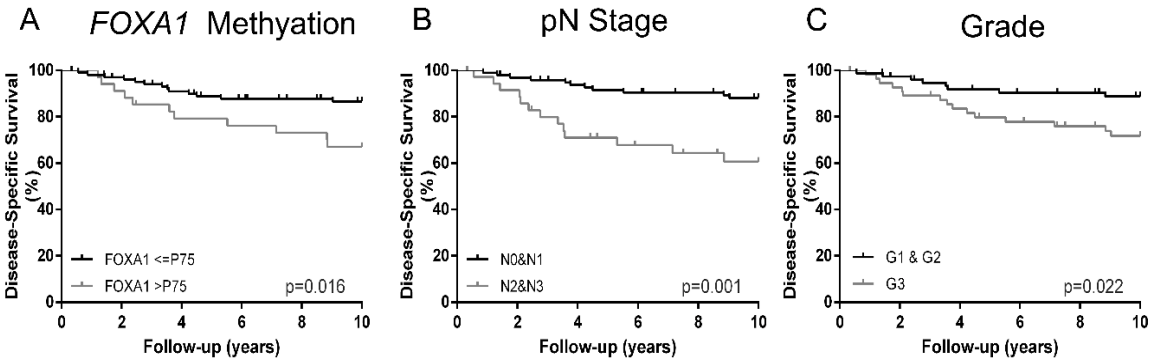

Figure 15 – Disease-specific survival curves (Kaplan–Meier with log rank test) of *FOXA1* methylation (A) and clinicopathological parameters [pN Stage (B) and Grade (C)]. Abbreviations – P75 – percentile 75 of methylation of *FOXA1*.

Moreover, pN2 or pN3 patients disclosed worse disease-specific survival than pN0 or pN1 patients (Figure 15 B). Similarly, patients with high grade BrC (G3) displayed a worse disease-specific survival than patients with low or moderate grade BrC (G1 or G2) (Figure 15 C).

A Cox regression analysis was also computed to assess the potential of both clinicopathological and epigenetic variables in predicting disease-specific survival (Table

10). In this multivariable model, only *FOXA1* methylation levels and pN Stage retained statistical significance for predicting disease-specific survival.

Table 10 – Cox regression models assessing the potential of clinical and epigenetic variables in the prediction of disease-specific survival for 127 patients with Breast Cancer.

Disease-Specific Survival	Variable	HR	CI (95%)	*p* value
Univariable	**Grade**			
	G1	1		
	G2 & G3	2.725	1.155 – 6.428	**0.022**
	pN Stage			
	N0 & N1	1		
	N2 &N3	4.061	1.814 – 9.089	**0.001**
	FOXA1			
	≤P75	1		
	>P75	2.678	1.200 – 5.978	**0.016**
Multivariable	**Grade**			
	G1	1		
	G2 & G3	2.005	0.082 – 4.866	0.124
	pN Stage			
	N0 & N1	1		
	N2 &N3	4.855	1.981 – 10.611	**<0.001**
	FOXA1			
	≤P75	1		
	>P75	2.710	1.161 – 6.324	**0.021**

HR – Hazard Ratio; CI – Confidence Interval; P75 – Percentile 75 of methylation levels of *FOXA1*

To disclose the potential of *FOXA1* methylation levels as predictors of disease-specific survival for each pN Stage category, a stratified disease-specific survival analysis was performed (Table 11). Thus, in pN0&pN1 group, patients with high *FOXA1* promoter methylation levels had a 3.755- fold increased risk of dying from BrC.

Table 11 – Cox regression model assessing the potential *FOXA1* methylation levels in a stratified analysis by pN Stage categories in the prediction of disease-specific survival for 132 patients with Breast Cancer.

Disease-Specific Survival				
Layering Variable	Variable	HR	CI (95%)	*p* value
pN0 & pN1	***FOXA1***			
	≤P75	1		
	>P75	3.755	1.091 – 11.722	**0.035**
pN2 & pN3	***FOXA1***			
	≤P75	1		
	>P75	2.668	0.868 – 8.201	0.087

HR – Hazard Ratio; CI – Confidence Interval; P75 – Percentile 75 of methylation levels of *FOXA1*

DISCUSSION

DISCUSSION

At global level, BrC remains the major cause of cancer-related death in women. Mammographic-based screening led to an increase in the incidence of BrC diagnosis and a concomitant reduction in mortality between 28 and 45 % in countries in which it was implemented [10, 13, 64, 65]. Mammography is able to detect BrC with 70% sensitivity and 92% specificity [66]. However, the mammography-based screening is dependent of radiologist experience and the accuracy of diagnosis is influenced by breast density [66]. Moreover, the benefit of mammographic BrC screening in women between 40-49 years is still a matter of controversy [64] and overdiagnosis (i.e., “detection of cancers that would never have been found if it was not for the screening test” and includes all cancers, in situ or invasive [65]) due to mammography screening is becoming an important issue in the scientific community. Efforts to find new diagnosis biomarkers with sensitivity and specificity superior to those of mammography and that might allow for stratification of BrC aggressiveness is an ongoing field of research. Thus, the purpose of this study was to assess the potential of DNA methylation as diagnostic and prognostic marker for BrC.

Currently, DNA methylation is acknowledged as an early event in cancer development, including BrC [55, 67, 68]. Several tumor suppressor genes have been found hypermethylated in precursor lesions of invasive carcinoma indicating that DNA methylation is an early event in breast carcinogenesis [55, 68-70]. Indeed, aberrant DNA methylation has gained recognition as a cancer-associated event, and the definition of tumor-specific methylome has become the focus of multiple studies [67]. Indeed, DNA methylation has been proposed as a valuable cancer biomarker owing to its link to tissue-specific gene silencing [71, 72].

In this dissertation, we showed that five out of the six candidate genes might be suitable markers for BrC diagnosis/detection (*APC, CCND2, FOXA1, RASSF1A* and *SCGB3A1*). These genes were able to accurately discriminate NrB from invasive BrC tissues, confirming in tissue samples the diagnostic potential of *APC, CCND2, RASSF1A* and *SCGB3A1* methylation as diagnosis marker that we previously reported for fine-needle aspirate washings of breast lesions [45, 50]. Remarkably, *FOXA1* promoter also demonstrated significant differences in methylation levels between BrC and NrB samples. Although there are multiple studies associating *FOXA1* expression and BrC prognosis and subtyping, this is the first study, to the best of our knowledge, that unveiled the potential of *FOXA1* methylation as diagnostic BrC marker. Conversely, in accordance with a previous publication, *BRCA1* methylation did not discriminate BrC from NBr tissue in our sample set [73]. This finding is also in line with a previous study from our research team that found no

significant difference in *BRCA1* methylation levels among BrC, NrB and fibroadenoma samples obtained through FNB [45].

Moreover, we sought to define a small panel of methylated genes that might accurately detect BrC, maximizing sensitivity and specificity. However, a low NPV was obtained, probably due to the small number of NBr samples. To overcome this limitation, we combined a highly sensitive methylation marker (*RASSF1A*) with highly specific methylation markers (*APC* and *FOXA1*) thus accomplishing a panel which demonstrated sensitivity, specificity and accuracy over 94%, as well as high PPV. *SCGB3A1* and *CCND2* methylation were not considered for this panel because they were not able to detect TNBC (data not shown). ROC curve analysis confirmed the good performance of the selected panel, with an AUC above 0.95. Although *FOXA1* methylation has not been previously reported as a diagnostic marker, *RASSF1A* and *APC* methylation were already reported as part of methylation panel developed for FNB samples [45, 50]. The initial study reported a four-gene panel for detection of BrC in FNB (*APC, CCND2, RASSF1A* and *SCGB3A1*) although due to the scarcity of DNA the panel was reduced to three-genes: *APC, CCND2* and *RASSF1A*) [45, 50]. The different panels proposed might be explained by differences among populations or the type of biological samples (tissue and body fluids) [49, 57, 74, 75]. Thus, large multicenter prospective studies are critical to validate the clinical use of the three-gene panel proposed herein.

Our second aim was to associate the six genes' promoter methylation levels with BrC molecular subtypes defined by IHC, following the updated ESMO guidelines for molecular classification (Table 3). Nevertheless, no associations were found between promoter methylation levels and BrC molecular subtypes. IHC has some limitations for the classification of TNBC/basal tumors and discrimination between luminal subtypes A and B [21, 30, 31]. Moreover, our cohort is composed of consecutive samples from BrC patients diagnosed between 1996 and 2001 whose molecular subtype frequencies were similar to those expected. Thus, the number of ERBB2 and TNBC tumor subtypes is rather limited, although it parallels the lower frequencies commonly found in a consecutive cohort of BrC patients. In fact, studies reporting associations between DNA methylation and specific molecular subtypes have used a similar frequency of all the subtypes or have only analyzed a specific subtype [43, 47, 48].

Furthermore, although IHC subtyping is more feasible and less expensive, those studies have used several microarrays platforms. Additionally, analysis of aberrant DNA methylation was also performed using different methods. Indeed, using array-based and whole genome-analysis for DNA methylation, a correlation between DNA methylation pattern and molecular subtypes was found, revealing BrC subsets with different prognosis within each BrC molecular subtype [44, 76-79]. Therefore, differences in molecular

subtypes' classification and IHC cut-off definitions as well as in aberrant methylation analyses prevents definitive comparisons among studies.

Concerning our last aim, i.e., the prognostic significance of gene-specific aberrant promoter methylation in BrC, the association between genes' promoter methylation and standard clinicopathological parameters (histological subtype, grade, hormone receptor status, ERBB2 status, pathological T and N stage and stage) was assessed for all candidate genes. Higher *APC* and *SCGB3A1* promoter methylation levels significantly associated with histological subtype. Generally, SSC showed lower *APC* methylation levels, but a significant difference was only depicted between SSC and ILC. Moreover, SSC displayed lower *SCGB3A1* methylation levels than any other histological subtype. Due to SSC heterogeneity and low frequency most of the studies do not associate methylation levels with SSC. However, Tisserand and colleagues reported the lack of *SCGB3A1* methylation in medullary carcinoma of the breast, a SSC type also included in TNBC [80].

Furthermore, in our study, hormone receptor status, an important prognostic marker, associated with *RASSF1A* and *FOXA1* methylation levels. Indeed, ER+/PR+ BrC tumors displayed higher *RASSF1A* methylation levels than ER+ BrC. Although, *RASSF1A* methylation has been shown to be an ER status discriminator [43, 56], we did not find differences considering only ER status – luminal vs. non-luminal BrC samples (data not shown), in accordance with previous reports [60, 81]. Moreover, most of the studies categorize hormone receptor status only into two categories: positive (positive for ER and/or PR) and negative. However, it is currently accepted that ER positive/PR positive and ER positive/PR negative BrC have different prognosis [9, 16]. Contrarily to *RASSF1A, FOXA1* methylation levels were significantly lower in ER+/PR+ BrC. Although, *FOXA1* methylation is still poorly studied in BrC, *FOXA1* hypermethylation was described in basal BrC cell lines, a negative hormone receptor subtype [52, 53].

Due to high mortality BrC rate one of our aims was also to assess the prognostic value of the selected aberrantly methylated genes in BrC. Thus, concerning disease-free survival only standard clinicopathological parameters were associated with survival in univariable analysis. Nevertheless, regarding disease-specific survival, *FOXA1* methylation, grade and pN stage were associated with patient outcome.

Although, ER and PR expression were previously associated with favorable prognosis and predicted response to endocrine therapy [9, 16], in our cohort hormone receptor status did not predict outcome. Because no selection bias was apparent in our series, this lack of prognostic value might be associated with the determination of hormone receptor status at diagnosis (some case were determined by pharmacology) leading to discrepancies to the results obtained by IHC and may have led to different therapeutic strategies. Remarkably, high *FOXA1* methylation levels (above percentile 75) associated

with shorter disease-specific survival. The cut-off value was based in previous studies performed in FNB from BrC patients that associated high *RASSF1A* methylation levels with shorter disease-free survival [50]. Empirically, the association of promoter methylation levels and gene expression correlates with the density of CpG dinucleotide methylation. Therefore, higher promoter methylation levels are more likely to cause a decrease in mRNA expression and, thus, a more effective gene silencing. Although *FOXA1* methylation levels have not been previously reported as prognostic parameters in BrC, *FOXA1* expression was previously associated with good prognosis and response to endocrine therapy [82, 83]. Therefore, *FOXA1* promoter methylation might be the mechanism underlying *FOXA1* downregulation that was associated with BrC carrying worse prognosis.

Multivariable analysis showed that the prognostic value of *FOXA1* methylation was dependent of pN stage. Indeed, *FOXA1* methylation prognostic value was limited to pN0 & pN1 group, which might be explained by the small number of cases in this group (n=36). Further studies are needed to better understand the potential of *FOXA1* methylation in BrC prognostication.

Moreover, in this series of BrC patients, we were not able to verify the prognostic value of *RASSF1A* methylation levels. Albeit *RASSF1A* hypermethylation was reported to be a poor prognostic marker in BrC, associated with shorter disease-free survival and shorter disease-specific survival [50, 84, 85], other studies had results similar to ours [86-88]. These differences might be due to variations among patient populations and methodologies used for assessment of methylation levels. However, a meta-analysis of published data suggests that *RASSF1A* methylation is, indeed, associated with poor prognosis both for disease-free and specific-disease survival [89]. Thus, larger prospective studies are still needed to further establish the clinical utility of *RASSF1A* methylation in BrC.

CONCLUSIONS AND FUTURE PERSPECTIVES

CONCLUSIONS AND FUTURE PERSPECTIVES

In this study we confirmed the value of DNA methylation as BrC diagnostic marker, as previously reported by us in FNB of BrC patients. Indeed, a three-gene panel (*APC, FOXA1* and *RASSF1A*) demonstrated high specificity, sensitivity and accuracy for BrC diagnosis. As future perspective, we expect to validate this panel as diagnostic marker in plasma samples. The potential of liquid biopsies, including blood samples, for cancer diagnosis / early detection has been the focus of many recent studies owing to its less invasive nature. These samples are enriched in circulating cell-free DNA (cfDNA) which is not associated with any cell fraction [90]. Several studies reported an increase of cfDNA in cancer patients and this has been validated as tumor biomarker in those patients [91, 92]. Moreover, detection of DNA-methylation in serum and plasma was already reported for diagnostic, monitoring and prognostic purposes in BrC patients, although with suboptimal performance [49, 57, 93, 94]. Thus, we will explore the possibility of using a DNA-methylation based test as a diagnosis marker with improved sensitivity and specificity.

Contrarily to our initial expectation, no correlation was found between DNA methylation and BrC molecular subtype. Because this might be due to the sample size, we intend to enrich the patient cohort with ERBB2 overexpressing and TNBC subtypes.

Moreover, we revealed the potential of *FOXA1* methylation as prognostic marker. High *FOXA1* methylation levels associated with shorter disease-specific survival dependent on pN stage. Thus, we intend to increase the number of pN2&pN3 tumors to verify whether *FOXA1* methylation levels might have prognostic value at more advanced stages. Additionally, to verify whether methylation is the main silencing mechanism of *FOXA1,* we will determine whether *FOXA1* methylation levels correlate inversely with mRNA expression levels.

Further studies, preferably involving multiple institutions, are required to further validate these findings. *FOXA1* promoter methylation might constitute an important prognostic ancillary tool for assisting clinical decision making.

REFERENCES

1. Ferlay, J., et al., Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer, 2015. **136**(5): p. E359-86.
2. DeSantis, C.E., et al., International Variation in Female Breast Cancer Incidence and Mortality Rates. Cancer Epidemiol Biomarkers Prev, 2015. **24**(10): p. 1495-506.
3. Ferlay, J., I. Soerjomataram, and M. Ervik, GLOBOCAN, cancer incidence and mortality worldwide: IARC cancer base no. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. 2012.
4. Ferlay, J., et al., Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer, 2013. **49**(6): p. 1374-403.
5. Burstein, H.J., J.R. Harris, and M. Morrow, Malignant tumors of the breast, in DeVita, Hellman, and Rosenberg's cancer : principles & practice of oncology, V.T. DeVita, T.S. Lawrence, and S.A. Rosenberg, Editors. 2011, Lippincott Williams & Wilkins. p. 1401-1446.
6. Vogel, V.G., Epidemiology of Breast Cancer, in Breast Pathology, D.J. Dabbs, Editor. 2012, Elsevier Health Sciences. p. 44-56.
7. Howlader, N., et al., SEER Cancer Statistics Review, 1975-2013.[Based on the November 2015 SEER data submission, posted to the SEER web site, April 2016.], in Bethesda, MD: National Cancer Institute. 2016.
8. DeSantis, C.E., et al., Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. CA: A Cancer Journal for Clinicians, 2016. **66**(1): p. 31-42.
9. Lakhani, S.R., WHO classification of tumours of the breast. 2012: International Agency for Research on Cancer.
10. Schuur, E.R. and J.P. DeAndrade, Breast Cancer: Molecular Mechanisms, Diagnosis, and Treatment, in International Manual of Oncology Practice: (iMOP) - Principles of Medical Oncology, A.R. de Mello, Á. Tavares, and G. Mountzios, Editors. 2015, Springer International Publishing: Cham. p. 155-200.
11. Senkus, E., et al., Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology, 2015. **26**(suppl 5): p. v8-v30.
12. Tuzlali, S., Pathology of Breast Cancer, in Breast Disease: Diagnosis and Pathology, A. Aydiner, A. İğci, and A. Soran, Editors. 2016, Springer International Publishing: Cham. p. 241-266.
13. Schnitt, S.J. and S.R. Lakhani, Breast Cancer, in World Cancer Report 2014, B. Stewart and C.P. Wild, Editors. 2014, International Agency for Research on Cancer: Lyon. p. 362-373.

14. Wu, Y. and A.A. Sahin, Prognostic and Predictive Factors of Invasive Breast Cancer, in Breast Disease: Diagnosis and Pathology, A. Aydiner, A. İğci, and A. Soran, Editors. 2016, Springer International Publishing: Cham. p. 187-206.
15. Edge SB, et al., eds. AJCC cancer staging manual. 7th ed. Vol. 304. 2010, Springer: New York. 364-375.
16. Dai, X., et al., Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes. J Cancer, 2016. **7**(10): p. 1281-94.
17. Győrffy, B., et al., Multigene prognostic tests in breast cancer: past, present, future. Breast Cancer Research, 2015. **17**(1): p. 11.
18. Harris, L.N., et al., Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice guideline. Journal of Clinical Oncology, 2016: p. JCO652289.
19. Network, N.C.C., NCCN Clinical Practice Guidelines–Breast Cancer. V. 2.2016. National Comprehensive Cancer Network Web site. 2016.
20. Nounou, M.I., et al., Breast Cancer: Conventional Diagnosis and Treatment Modalities and Recent Patents and Technologies. Breast cancer: basic and clinical research, 2015. **9**(Suppl 2): p. 17.
21. Coates, A.S., et al., Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Annals of Oncology, 2015: p. mdv221.
22. Dogan, A.L., Biology and Genetics of Breast Cancer, in Breast Disease: Diagnosis and Pathology, A. Aydiner, A. İğci, and A. Soran, Editors. 2016, Springer International Publishing: Cham. p. 145-160.
23. Hirshfield, K.M., T.R. Rebbeck, and A.J. Levine, Germline mutations and polymorphisms in the origins of cancers in women. Journal of oncology, 2010. **2010**.
24. Perou, C.M., et al., Molecular portraits of human breast tumours. Nature, 2000. **406**(6797): p. 747-752.
25. Sørlie, T., et al., Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences, 2001. **98**(19): p. 10869-10874.
26. Sørlie, T., et al., Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences, 2003. **100**(14): p. 8418-8423.
27. van 't Veer, L.J., et al., Gene expression profiling predicts clinical outcome of breast cancer. Nature, 2002. **415**(6871): p. 530-536.

28. Prat, A., et al., Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. The oncologist, 2013. **18**(2): p. 123-133.
29. Prat, A., et al., Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Research, 2010. **12**(5): p. 1-18.
30. Maisonneuve, P., et al., Proposed new clinicopathological surrogate definitions of luminal A and luminal B (HER2-negative) intrinsic breast cancer subtypes. Breast Cancer Research, 2014. **16**(3): p. 1-9.
31. Prat, A., et al., Prognostic Significance of Progesterone Receptor–Positive Tumor Cells Within Immunohistochemically Defined Luminal A Breast Cancer. Journal of Clinical Oncology, 2012.
32. García-Giménez, J.L., T. Ushijima, and T.O. Tollefsbol, Chapter 1 - Epigenetic Biomarkers: New Findings, Perspectives, and Future Directions in Diagnostics, in Epigenetic Biomarkers and Diagnostics. 2016, Academic Press: Boston. p. 1-18.
33. Hagood, J.S., Beyond the Genome: Epigenetic Mechanisms in Lung Remodeling. Physiology, 2014. **29**(3): p. 177-185.
34. Kanwal, R., K. Gupta, and S. Gupta, Cancer Epigenetics: An Introduction, in Cancer Epigenetics: Risk Assessment, Diagnosis, Treatment, and Prognosis, M. Verma, Editor. 2015, Springer New York: New York, NY. p. 3-25.
35. Basse, C. and M. Arock, The increasing roles of epigenetics in breast cancer: Implications for pathogenicity, biomarkers, prevention and treatment. Int J Cancer, 2014.
36. Sharma, S., T.K. Kelly, and P.A. Jones, Epigenetics in cancer. Carcinogenesis, 2010. **31**(1): p. 27-36.
37. Portela, A. and M. Esteller, Epigenetic modifications and human disease. Nat Biotech, 2010. **28**(10): p. 1057-1068.
38. Xie, H., Chapter 5 - Personalized Epigenetics: Analysis and Interpretation of DNA Methylation Variation A2 - Tollefsbol, Trygve O, in Personalized Epigenetics. 2015, Academic Press: Boston. p. 123-150.
39. Cervera, R., et al., Chapter 15 - DNA Methylation in Breast Cancer A2 - García-Giménez, José Luis, in Epigenetic Biomarkers and Diagnostics. 2016, Academic Press: Boston. p. 297-312.
40. Costa-Pinheiro, P., et al., Diagnostic and prognostic epigenetic biomarkers in cancer. Epigenomics, 2015(0).
41. Győrffy, B., et al., Aberrant DNA methylation impacts gene expression and prognosis in breast cancer subtypes. International Journal of Cancer, 2016. **138**(1): p. 87-97.

42. Hoque, M.O., et al., Detection of aberrant methylation of four genes in plasma DNA for the detection of breast cancer. J Clin Oncol, 2006. **24**(26): p. 4262-9.
43. Sunami, E., et al., Estrogen receptor and HER2/neu status affect epigenetic differences of tumor-related genes in primary breast tumors. Breast Cancer Res, 2008. **10**(3): p. R46.
44. Holm, K., et al., Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. Breast Cancer Res, 2010. **12**(3): p. R36.
45. Jeronimo, C., et al., Quantitative hypermethylation of a small panel of genes augments the diagnostic accuracy in fine-needle aspirate washings of breast lesions. Breast Cancer Res Treat, 2008. **109**(1): p. 27-34.
46. Liu, L., et al., Quantitative detection of methylation of FHIT and BRCA1 promoters in the serum of ductal breast cancer patients. Bio-Medical Materials and Engineering, 2015. **26**(s1): p. S2217-S2222.
47. Stefansson, O.A., et al., CpG island hypermethylation of BRCA1 and loss of pRb as co-occurring events in basal/triple-negative breast cancer. Epigenetics, 2011. **6**(5): p. 638-649.
48. Lee, J.S., et al., Basal-like breast cancer displays distinct patterns of promoter methylation. Cancer biology & therapy, 2010. **9**(12): p. 1017-1024.
49. Radpour, R., et al., Hypermethylation of tumor suppressor genes involved in critical regulatory pathways for developing a blood-based test in breast cancer. PLoS One, 2011. **6**(1): p. e16080.
50. Martins, A.T., et al., High RASSF1A promoter methylation levels are predictive of poor prognosis in fine-needle aspirate washings of breast cancer lesions. Breast Cancer Res Treat, 2011. **129**(1): p. 1-9.
51. Ghosh, S., et al., Genome-wide DNA methylation profiling reveals parity-associated hypermethylation of FOXA1. Breast Cancer Res Treat, 2014. **147**(3): p. 653-9.
52. Gong, C., et al., FOXA1 repression is associated with loss of BRCA1 and increased promoter methylation and chromatin silencing in breast cancer. Oncogene, 2014.
53. Locke, W.J., et al., Coordinated epigenetic remodelling of transcriptional networks occurs during early breast carcinogenesis. Clin Epigenetics, 2015. **7**(1): p. 52.
54. Donninger, H., M.D. Vos, and G.J. Clark, The RASSF1A tumor suppressor. Journal of cell science, 2007. **120**(18): p. 3163-3172.
55. van Hoesel, A.Q., et al., Assessment of DNA methylation status in early stages of breast cancer development. Br J Cancer, 2013. **108**(10): p. 2033-8.

56. Kajabova, V., et al., RASSF1A Promoter Methylation Levels Positively Correlate with Estrogen Receptor Expression in Breast Cancer Patients. Transl Oncol, 2013. **6**(3): p. 297-304.
57. Kloten, V., et al., Promoter hypermethylation of the tumor-suppressor genes ITIH5, DKK3, and RASSF1A as novel biomarkers for blood-based breast cancer screening. Breast Cancer Research, 2013. **15**(1): p. R4.
58. Krop, I., et al., HIN-1, an inhibitor of cell growth, invasion, and AKT activation. Cancer research, 2005. **65**(21): p. 9659-9669.
59. Feng, W., et al., Correlation between CpG methylation profiles and hormone receptor status in breast cancers. Breast Cancer Res, 2007. **9**(4): p. R57.
60. Benevolenskaya, E.V., et al., DNA methylation and hormone receptor status in breast cancer. Clinical epigenetics, 2016. **8**(1): p. 1.
61. Clark, S.J., et al., DNA methylation: bisulphite modification and analysis. Nat Protoc, 2006. **1**(5): p. 2353-64.
62. Zhao, F. and B. Bapat, Chapter 8 - The Role of Methylation-Specific PCR and Associated Techniques in Clinical Diagnostics A2 - García-Giménez, José Luis, in Epigenetic Biomarkers and Diagnostics. 2016, Academic Press: Boston. p. 155-173.
63. Shames, D.S., J.D. Minna, and A.F. Gazdar, Methods for detecting DNA methylation in tumors: from bench to bedside. Cancer Lett, 2007. **251**(2): p. 187-98.
64. Warner , E., Breast-Cancer Screening. New England Journal of Medicine, 2011. **365**(11): p. 1025-1032.
65. Independent, U.K.P.o.B.C.S., The benefits and harms of breast cancer screening: an independent review. The Lancet, 2012. **380**(9855): p. 1778-1786.
66. Pisano, E.D., et al., Diagnostic performance of digital versus film mammography for breast-cancer screening. New England Journal of Medicine, 2005. **353**(17): p. 1773-1783.
67. Esteller, M., Epigenetics in cancer. N Engl J Med, 2008. **358**(11): p. 1148-59.
68. Widschwendter, M. and P.A. Jones, DNA methylation and breast carcinogenesis. Oncogene, 2002. **21**(35): p. 5462-5482.
69. Muggerud, A.A., et al., Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. Breast Cancer Research, 2010. **12**(1): p. 1.
70. Fackler, M.J., et al., DNA methylation of RASSF1A, HIN-1, RAR-β, Cyclin D2 and Twist in in situ and invasive lobular breast carcinoma. International journal of cancer, 2003. **107**(6): p. 970-975.

71. Heyn, H. and M. Esteller, DNA methylation profiling in the clinic: applications and challenges. Nat Rev Genet, 2012. **13**(10): p. 679-92.
72. Jeronimo, C. and R. Henrique, Epigenetic biomarkers in urological tumors: A systematic review. Cancer Lett, 2014. **342**(2): p. 264-74.
73. Li, Z., et al., Methylation profiling of 48 candidate genes in tumor and matched normal tissues from breast cancer patients. Breast cancer research and treatment, 2015. **149**(3): p. 767-779.
74. Shan, M., et al., Detection of aberrant methylation of a six-gene panel in serum DNA for diagnosis of breast cancer. Oncotarget, 2016.
75. Fackler, M.J., et al., Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer. Cancer Res, 2004. **64**(13): p. 4442-52.
76. Bediaga, N.G., et al., DNA methylation epigenotypes in breast cancer molecular subtypes. Breast Cancer Research, 2010. **12**(5): p. 1.
77. Stirzaker, C., E. Zotenko, and S.J. Clark, Genome-wide DNA methylation profiling in triple-negative breast cancer reveals epigenetic signatures with important clinical value. Molecular & Cellular Oncology, 2016. **3**(1): p. e1038424.
78. Stefansson, O.A., et al., A DNA methylation-based definition of biologically distinct breast cancer subtypes. Mol Oncol, 2015. **9**(3): p. 555-68.
79. Cancer Genome Atlas, N., Comprehensive molecular portraits of human breast tumours. Nature, 2012. **490**(7418): p. 61-70.
80. Tisserand, P., et al., Lack of HIN-1 methylation defines specific breast tumor subtypes including medullary carcinoma of the breast and BRCA1-linked tumors. Cancer biology & therapy, 2003. **2**(5): p. 559-563.
81. Widschwendter, M., et al., Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. Cancer research, 2004. **64**(11): p. 3807-3813.
82. Mehta, R.J., et al., FOXA1 is an independent prognostic marker for ER-positive breast cancer. Breast cancer research and treatment, 2012. **131**(3): p. 881-890.
83. Albergaria, A., et al., Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. Breast Cancer Research, 2009. **11**(3): p. 1.
84. Buhmeida, A., et al., RASSF1A methylation is predictive of poor prognosis in female breast cancer in a background of overall low methylation frequency. Anticancer research, 2011. **31**(9): p. 2975-2981.
85. Muller, H.M., et al., DNA methylation in serum of breast cancer patients: an independent prognostic marker. Cancer Res, 2003. **63**(22): p. 7641-5.

86. Klajic, J., et al., Quantitative DNA methylation analyses reveal stage dependent DNA methylation and association to clinico-pathological factors in breast tumors. BMC Cancer, 2013. **13**.
87. Sharma, G., et al., Prognostic relevance of promoter hypermethylation of multiple genes in breast cancer patients. Analytical Cellular Pathology, 2009. **31**(6): p. 487-500.
88. Cho, Y.H., et al., Prognostic significance of gene-specific promoter hypermethylation in breast cancer patients. Breast cancer research and treatment, 2012. **131**(1): p. 197-205.
89. Jiang, Y., et al., The prognostic role of RASSF1A promoter methylation in breast cancer: a meta-analysis of published data. PLoS One, 2012. **7**(5): p. e36780.
90. Warton, K. and G. Samimi, Methylation of cell-free circulating DNA in the diagnosis of cancer. Frontiers in Molecular Biosciences, 2015. **2**.
91. Marzese, D.M., H. Hirose, and D.S. Hoon, Diagnostic and prognostic value of circulating tumor-related DNA in cancer patients. Expert Rev Mol Diagn, 2013. **13**(8): p. 827-44.
92. Schwarzenbach, H., D.S.B. Hoon, and K. Pantel, Cell-free nucleic acids as biomarkers in cancer patients. Nature Reviews Cancer, 2011. **11**(6): p. 426-437.
93. Gobel, G., et al., Prognostic significance of methylated RASSF1A and PITX2 genes in blood- and bone marrow plasma of breast cancer patients. Breast Cancer Res Treat, 2011. **130**(1): p. 109-17.
94. Avraham, A., et al., Serum DNA methylation for monitoring response to neoadjuvant chemotherapy in breast cancer patients. International Journal of Cancer, 2012. **131**(7): p. E1166-E1172.

SUPPLEMENTARY MATERIAL

Appendix I – Anatomic Stage/Prognostic Groups adapted from [15]

Stage 0	**Tis**	**N0**	**M0**
Stage IA	T1	N0	M0
Stage IB	T0	N1mi	M0
	T1	N1mi	M0
Stage IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0, T1, T2	N2	M0
	T3	N1, N2	M0
	T4	N0, N1, N2	M0
Stage IIIB	T4	N0, N1, N2	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

Appendix II –TNM classification of carcinomas of the breast adapted from [15]

Primary tumor (T)*	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ
Tis (DCIS)	Ductal carcinoma in situ
Tis (LCIS)	Lobular carcinoma in situ
Tis (Paget's)	Paget's disease of the nipple NOT associated with invasive carcinoma and/or carcinoma in situ (DCIS and/or LCIS) in the underlying breast parenchyma. Carcinomas in the breast parenchyma associated with Paget's disease are categorized based on the size and characteristics of the parenchymal disease, although the presence of Paget's disease should still be noted
T1	Tumor ≤20 mm in greatest dimension
T1mi	Tumor ≤1 mm in greatest dimension
T1a	Tumor >1 mm but ≤5 mm in greatest dimension
T1b	Tumor >5 mm but ≤10 mm in greatest dimension
T1c	Tumor >10 mm but ≤20 mm in greatest dimension
T2	Tumor >20 mm but ≤50 mm in greatest dimension
T3	Tumor >50 mm in greatest dimension
T4	Tumor of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules). *Note*: Invasion of the dermis alone does not qualify as T4
T4a	Extension to the chest wall, not including only pectoralis muscle adherence/invasion
T4b	Ulceration and/or ipsilateral satellite nodules and/or edema (including peau d'orange) of the skin, which do not meet the criteria for inflammatory carcinoma
T4c	Both T4a and T4b
T4d	Inflammatory carcinoma
Regional Lymph Nodes (N) – Clinical	
Nx	Regional lymph nodes cannot be assessed (e.g previously removed)
N0	No regional lymph node metástases
N1	Metastases to movable ipsilateral level I, II axillary lymph node(s)

N2	Metastases in ipsilateral level I, II axillary lymph nodes that are clinically fixed or matted; or in clinically detected ** ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastases
N2a	Metastases in ipsilateral level I, II axillary lymph nodes fixed to one another (matted) or to other structures
N2b	Metastases only in clinically detected** ipsilateral internal mammary nodes and in the absence of clinically evident level I, II axillary lymph node metastases
N3	Metastases in ipsilateral infraclavicular (level III axillary) lymph node(s) with or without level I, II axillary lymph node involvement; or in clinically detected** ipsilateral internal mammary lymph node(s) with clinically evident level I, II axillary lymph node metastases; or metastases in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement
N3a	Metastases in ipsilateral infraclavicular lymph node(s)
N3b	Metastases in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)
N3c	Metastases in ipsilateral supraclavicular lymph node(s)
Regional Lymph Nodes (N) – Pathological	
pNX	Regional lymph nodes cannot be assessed (e.g., previously removed, or not removed for pathologic study)
pN0	No regional lymph node metastasis identified histologically
pN1	Micrometastases; or metastases in 1–3 axillary lymph nodes; and/or in internal mammary nodes with metastases detected by sentinel lymph node biopsy but not clinically detected **
pN1mi	Micrometastases (greater than 0.2 mm and/or more than 200 cells, but none greater than 2.0 mm)
pN1a	Metastases in 1–3 axillary lymph nodes, at least one metastasis greater than 2.0 mm
pN1b	Metastases in internal mammary nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected **
pN1c	Metastases in 1–3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected***
pN2	Metastases in 4–9 axillary lymph nodes; or in clinically detected ** internal mammary lymph nodes in the absence of axillary lymph node metastases

pN2a	Metastases in 4–9 axillary lymph nodes (at least one tumor deposit greater than 2.0 mm)
pN2b	Metastases in clinically detected** internal mammary lymph nodes in the absence of axillary lymph node metastases
pN3	Metastases in ten or more axillary lymph nodes; or in infraclavicular (level III axillary) lymph nodes; or in clinically detected** ipsilateral internal mammary lymph nodes in the presence of one or more positive level I, II axillary lymph nodes; or in more than three axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected** ; or in ipsilateral supraclavicular lymph nodes
pN3a	Metastases in ten or more axillary lymph nodes (at least one tumor deposit greater than 2.0 mm); or metastases to the infraclavicular (level III axillary lymph) nodes
pN3b	Metastases in clinically detected *** ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary lymph nodes; or in more than three axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected ***
pN3c	Metastases in ipsilateral supraclavicular lymph nodes
Metastasis (M)* **	
M0	No clinical or radiographic evidence of distant metastases
M1	Distant detectable metastases as determined by classic clinical and radiographic means and/or histologically proven larger than 0.2 mm

*T classification is the same regardless of whether it is based on clinical or pathological criteria; ** “Clinically detected” is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination and having characteristics highly suspicious for malignancy or a presumed pathologic macrometastasis based on fine needle aspiration biopsy with cytologic examination; ***M classification is the same regardless of whether it is based on clinical or pathological criteria.

Appendix III – Sodium bisulfite modification using EZ DNA Methylation-Gold™ Kit

Genomic DNA from all samples was modified by sodium bisulfite using EZ DNA Methylation-Gold™ Kit (Zymo Research, USA) according to the manufacturer's guidelines.

To sodium bisulfite modification, it was used 1µg of DNA obtained from fresh frozen sections. From each sample, DNA amount required was added to sterile distilled water up to a final volume of 20µL according to each sample concentration. In a PCR tube, 130µL of the CT Conversion Reagent was added to the samples DNA and incubated in The Applied Biosystems Veriti 96-Well Thermal Cycler (Thermo Fisher Scientific, USA) at 98ºC for 10 minutes, 64ºC for 3 hours.

Following the incubation, samples were transferred to a Zymo-Spin™ IC Column with 600µL of M-Binding Buffer. The columns were centrifuged at 10 000 rpm for 30 seconds, washed with 100 µL of M-Wash Buffer and centrifuged again. To each column was added 200µL of M-Desulphonation Buffer and let stand for 20 minutes at room temperature, followed by a centrifugation 10,000 rpm for 30 seconds to buffer discarded. The columns were washed twice with 200µL of M-Wash Buffer followed by a centrifugation at 10,000 rpm for 30 seconds.

At last, the columns were placed in 1.5mL microcentrifuge tube and DNA was eluted with 60µL of sterile distilled water for 5 minutes followed by a centrifugation at 13,000 rpm for 30 seconds. The modified DNA was stored at -80°C until further use.

One µg of CpGenome™ Universal Methylated DNA (Millipore, USA) was also modified, according to the method described above and eluted in 20µL of M-elution buffer.

Appendix IV – Empirical Cut-off for each gene based in ROC Curve analysis

Gene	Cut-off
APC	16.9889
CCND2	1.13523
FOXA1	30.6928
RASSF1A	62.2572
SCGB3A1	23.0525